Ranaghan, K. E., Shchepnovska, D., Bennie, S. J., Lawan, N., Macrae, S., Zurek, J., Manby, F. R., & Mulholland, A. J. (2019). Projector-Based Embedding Eliminates Density Functional Dependence for QM/MM Calculations of Reactions in Enzymes and Solution. *Journal of Chemical Information and Modeling*, *59*(5), 2063-2078. https://doi.org/10.1021/acs.jcim.8b00940

Computational Chemistry

# Projector-based embedding eliminates density functional dependence for QM/MM calculations of reactions in enzymes and solution

Kara Elizabeth Ranaghan, Darya Shchepnovska, Simon Jonathan Bennie, Narin Lawan, Stephen Macrae, Jolanta Zurek, Frederick R. Manby, and Adrian J. Mulholland

*J. Chem. Inf. Model.*, **Just Accepted Manuscript** • DOI: 10.1021/acs.jcim.8b00940 • Publication Date (Web): 22 Feb 2019

# Projector-based embedding eliminates density functional dependence for QM/MM calculations of reactions in enzymes and solution

Kara E. Ranaghan, Darya Shchepanovska, Simon J. Bennie, Narin Lawan[^], Stephen J. Macrae[§], Jolanta Zurek, Frederick R. Manby and Adrian J. Mulholland*

AUTHOR INFORMATION

Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol, UK BS8 1TS.

Corresponding Author

*Adrian J Mulholland: Adrian.Mulholland@bristol.ac.uk

Present Addresses

[^] Department of Chemistry, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand 50200.

[§]Southern Water, Worthing, West Sussex, UK.

KEYWORDS: Quantum mechanics / molecular mechanics (QM/MM), Chorismate mutase, enzyme catalysis, projector embedding.

ABSTRACT

Combined quantum mechanics/molecular mechanics (QM/MM) methods are increasingly widely utilized in studies of reactions in enzymes and other large systems. Here, we apply a range of QM/MM methods to investigate the Claisen rearrangement of chorismate to prephenate, in solution, and in the enzyme chorismate mutase. Using projector-based embedding in a QM/MM framework, we apply treatments up to the CCSD(T) level. We test a range of density functional QM/MM methods and QM region sizes. The results show that the calculated reaction energetics are significantly more sensitive to the choice of density functional than they are to the size of the QM region in these systems. Projector-based embedding of a wavefunction method in DFT reduced the 13 kcal/mol spread in barrier heights calculated at the DFT/MM level to a spread of just 0.3 kcal/mol, essentially eliminating dependence on the functional. Projector-based embedding of correlated *ab initio* methods provides a practical method for achieving high accuracy for energy profiles derived from DFT and DFT/MM calculations for reactions in condensed phases.

## INTRODUCTION

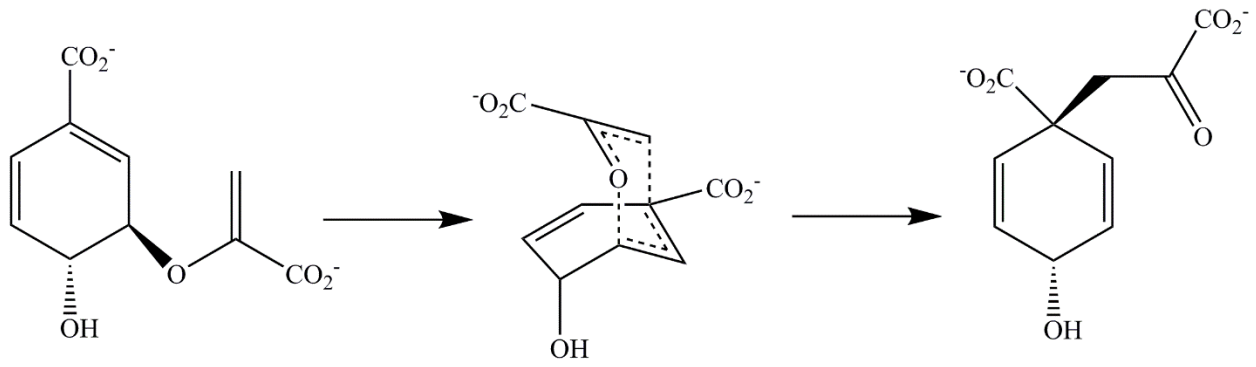
The award of the 2013 Nobel Prize for Chemistry to Martin Karplus, Michael Levitt and Arieh Warshel, recognised the development of multiscale methods and the important role these methods can now play in the understanding of biological systems.[1-3] Multiscale techniques have greatly benefited from recent advances in computer hardware and software and can provide insight into biological[4] and solid state systems.[5, 6] Hybrid quantum mechanics/molecular mechanics (QM/MM) technique are an exemplar of multiscale modelling methods. They combine an electronic structure treatment of a small region (e.g. the active site of an enzyme) with an empirical MM forcefield treatment of the bulk of the system. Most QM/MM studies apply semiempirical or density functioanl theory (DFT) methods for the QM part of the calculation. DFT methods potentially offer a good combination of accuracy and computational cost, but suffer from some well-known limitations and are not systematically improvable.[7, 8] When different density functionals give different results for the same system, it is not obvious which result should be preferred. DFT can be quite inaccurate for some applications, which may lead to qualitatively incorrect mechanistic conclusions[9-11] It is now possible to apply potentially highly accurate correlated *ab initio* electronic structure methods, e.g. coupled cluster, in QM/MM calculations. Such methods allow the calculation of activation barriers to within chemical accuracy (1 kcal/mol, considered the 'gold standard' of quantum chemical techniques)[11-13] necessary for reliable predictions of mechanism and detailed comparison with experiment. With the availability of such highly accurate *ab initio* methods in QM/MM calculations,[11-14] focus falls again on other aspects in achieving reliable predictions, for example, the appropriate choice of size of QM/MM region, and treatment of the boundary between QM and MM regions.[15] QM/MM methods can provide an accurate description of molecular interactions,[15, 16] and chemical properties, but of course involve approximations that should be tested (such as the limitations of empirical MM forcefields, and effects at the QM/MM boundary). It is possible to test consistency between QM and MM representations in various ways

e.g. by free energy calculations, calculating the difference between QM and QM/MM treatments of particular chemical species.[4, 17, 18] An obvious question in any QM/MM application is: what size should the QM and MM regions be? This has been investigated by e.g. the convergence of energetic properties with respect to QM region, with some suggesting that QM regions must be 100s of atoms in size.[19-23] QM/MM studies of catechol O-methyltransferase (COMT) provide a central example.[19, 23-25] Kulik *et al.* varied the QM region size from 64 atoms to 940 atoms to test the effect of the QM region size on the potential energy barrier in COMT. They suggested that 500-600 QM atoms are required for convergence of the activation barrier if the residues are chosen by distance alone and 200-300 QM atoms using a charge shift analysis to choose the residues in a more rigorous manner. In contrast, Jindal and Warshel[24] examined the effect of the size of the QM region on the free energy barrier to the reaction in COMT, and found little difference in activation free energy barrier; instead, they stressed the importance of adequate sampling rather than QM region size. Das *et al*.[26] have found rapid convergence with respect to QM region size for free energy profiles for proton transfer reactions in DNA, again stressing the importance of extensive sampling. Laio and Thiel[27] have shown that, while the activation energies for acetylene hydratase are quite slow to converge with respect to QM region size, improved accuracy does not change any mechanistic conclusions. They suggest using a minimal sized QM region to decide the mechanism and then probing the energetics further by enlarging the QM region. While it might initially be assumed that convergence can be guaranteed by enlarging the QM region, this is unlikely to be true with typical QM/MM calculations, particularly in asymmetric systems, or systems containing charges, for which errors associated with e.g. lack of charge transfer to, and polarization of, the MM region, particularly at the QM/MM boundary, are likely to increase with the size of the system and of the interface; some of these limitations of the QM/MM method will not converge as the QM region size increases. Use of frozen density approaches[28, 29] or similar methods may be appropriate for the QM/MM boundary, e.g. for consistency with an invariant charge MM model. Different QM/MM partitioning choices may be required, even for the same system, for calculations of different types of properties. When structural changes are involved, limitations of both the QM and MM methods could affect the results. An example is provided by Kaiyawet *et al*.[30], who investigated the mechanism of Michael

addition and covalent complex formation in thymidylate synthase. They tested the convergence of the QM/MM results with respect to QM region size and found that the inclusion of Arg116 was important as it interacts with the thiolate nucleophile. The increased accuracy with a larger QM region was due to limitations in the MM model of arginine which did not reproduce the energetics of distortion of the guanidinium side chain correctly, rather than failings of the QM/MM method.

A practical approach to accurate energy profiles for e.g. reactions in condensed phases (using QM or QM/MM methods) is through a projector-based approach for wavefunction (WF) in DFT embedding to obtain accurate potential energy surfaces for a reaction.[31] This has recently been used to calculate accurate QM/MM potential energy surfaces for the deprotonation of acetyl-coA in citrate synthase.[32, 33] The projector embedding approach allows the partitioning of regions in many types of system, making it relatively straightforward to select a small number of QM atoms for treatment with CC theory. The embedding of CC in DFT can overcome the limitations of commonly used density functionals, potentially removing the uncertainty the choice of functional and allowing for the calculation of chemical properties with high accuracy.[32, 34-37]

**Scheme 1.** The Claisen rearrangement of chorismate to form prephenate, the reaction catalyzed by the enzyme chorismate mutase.




Chorismate mutase (CM) catalyses the Claisen rearrangement of chorismate to prephenate

(Scheme 1) and has been a test case for understanding enzyme catalysis and the development of QM/MM methods for many years.[38-46] In part, this is driven by the relative simplicity of the Claisen rearrangement reaction (which occurs in the enzymes, and also, more slowly, in aqueous solution), and by the availability of experimental data. The free energy barrier $\Delta^{\ddagger}G$ = 15.4 kcal/mol ($\Delta^{\ddagger}H$ = 12.7 kcal/mol) derived from experimental kinetics for *Bacillus subtilis* CM (BsCM) at 25 °C is significantly lower than that for the uncatalysed reaction in aqueous solution ($\Delta^{\ddagger}G$ = 24.5 kcal/mol, $\Delta^{\ddagger}H$ = 20.7 kcal/mol).[47] This translates to a rate acceleration of $10^6$ by the enzyme ($\Delta\Delta^{\ddagger}G$ = 9.1 kcal/mol). An enthalpic barrier to reaction at 300 K of 13.1 (± 1.1) kcal/mol has been calculated for the reaction in BsCM using LCCSD(T0)/CHARMM27 QM/MM techniques for the potential energy barrier with zero-point energy and thermal corrections included at the DFT level.[12] This value is within chemical accuracy (i.e. 1 kcal/mol) of the experimental value of 12.7 kcal/mol. Linear scaling DFT methods have been used to optimise pathways (and associated transition state structures) for the CM reaction in large models of up to 2000 atoms, showing good convergence with QM region size.[48] These large-scale QM results (obtained using ONETEP) are in excellent agreement with those from QM/MM methods; in terms of barrier height (13.6 (±1.3) kcal/mol) and stabilization energies for active site residues. QM and QM/MM calculations (e.g. at the semiempirical[38, 40, 42, 49, 50], *ab initio* [51, 52] or DFT[53, 54] QM levels) have previously shown TS stabilization by the enzyme to be the major contribution to catalysis. It is encouraging that the importance of TS stabilization has been shown in the results of both QM and QM/MM calculations at all levels of theory, showing that QM/MM methods can provide valuable insight into this reaction. The electrostatic nature of catalysis is also supported by mutagenesis experiments[55, 56] and simulations[57, 58] that show a significant decrease in catalytic activity when Arg90 is mutated to the isosteric but neutral citrulline residue.

Here, we model the reaction in BsCM and in solution using QM/MM methods. We test a range of density functionals, and apply projector-based embedding[31, 32] for correlated *ab initio* calculations (up to the CCSD(T) level) on the reaction. The reaction energetics (barrier and reaction energy) differ greatly when different density functionals are used. The results are much more sensitive to the choice of density functional than they are to the size of the QM region. This

variation is removed when projector-based embeded *ab initio* methods are applied, yielding results in good agreement with experiment.

## METHODS

### *QM/MM Modelling of the reaction in the Enzyme and in Solution*

To model the reaction in the chorismate mutase enzyme, we constructed a model following procedures similar to those we have described previously.[12, 53, 54] Specifically, we used the structure of *Bacillus subtilis* chorismate mutase taken from the Protein Databank (PDB code 2CHT[59]), containing Bartlett's inhibitor[60] (a TS analogue) bound in the active site. Chorismate was substituted for the TS analogue as described in reference 55. This initial enzyme structure was hydrogenated and solvated (with a cut-off radius of 2.8 Å; the CHARMM[61] version of the TIP3P model[62] was used for water molecules) before being truncated to a sphere of radius 25 Å. The solution model was generated by solvating chorismate in a 25 Å sphere of CHARMM-type TIP3P[62] water molecules. Multiple structures were then generated by semi-empirical QM/MM molecular dynamics simulations of the BsCM complex and CM in solution (at the SCCDFTB[63, 64]/CHARMM22[65] level using stochastic boundary conditions) with the chorismate QM region restrained to be close to the TS (to a reaction coordinate value of $r = -0.5$ Å, where $r = d$(C2-O17) – $d$(C4-C14)/Å, see Figure 1 for atom labels). This reaction coordinate has been used in our[49, 51, 53, 54] and other previous studies[40]; in the current work, we tested it against nudged elastic band (NEB)[66, 67] calculations of the reaction pathway (see Supporting Information (SI)). Figure S1 shows a comparison of a potential energy profile calculated by adiabatic mapping with a profile generated using NEB techniques. The potential energy barrier is 12.2 kcal/mol with adiabatic mapping techniques and 11.5 kcal/mol in the CI-NEB profile. The shape of the adiabatic mapping profile agrees

well with the CI-NEB path and indicates that the reaction coordinate used here performs well for this reaction.

Structures were taken from the corresponding solution or enzyme SCCDFTB/CHARMM22 MD stochastic boundary simulations (120 ps) of the model systems for use in adiabatic mapping calculations. For this, Jaguar[68] and Tinker[69], linked by the interface program QoMMMa[70], were used for QM and MM calculations, with electronic coupling between the two regions treated by including MM charges in the QM Hamiltonian. CHARMM27 Lennard-Jones parameters (for standard CHARMM27 atom types, see Table S1) were used to describe QM/MM van der Waals interactions. The QM region was treated at the hybrid density functional B3LYP/6-31G(d) level of theory, which gives a reasonably good description of the reaction.[12] The enzyme model consisted of 7077 atoms and the solution model contained 7218 atoms. The MM region comprised an approximate 25 Å radius sphere of protein and/or solvent, treated with the CHARMM27 forcefield.[71] The outer 5 Å in each case was fixed (3324 atoms fixed in the enzyme model and 3535 fixed in the solution model), with all other atoms free to move. As there is no evidence for large-scale conformational changes during the reaction, this approach should give a representative sample of reactive conformations in the enzyme.[49, 51, 53, 54] Each initial structure was fully optimized at the B3LYP/6-31G(d)/CHARMM27 QM/MM level, while restraining the reaction coordinate ($r$) to –0.5 Å with a harmonic force constant of 500 kcal/mol/Å$^2$, to generate starting structures. Reaction pathways were generated by restrained optimizations in both directions along the reaction coordinate, towards the reactant and the product, in steps of 0.2 Å (0.1 Å around the TS), with both the MM and QM systems fully and consistently optimized at each step. Energy profiles were calculated from $r = -2.2$ Å to 2.2 Å, to identify the reactant and product minima. Reoptimization of the reactant complex without restraints, both in the enzyme and solution,

gave structures very similar to the lowest energy restrained structures (and energy differences less than 1 kcal/mol).

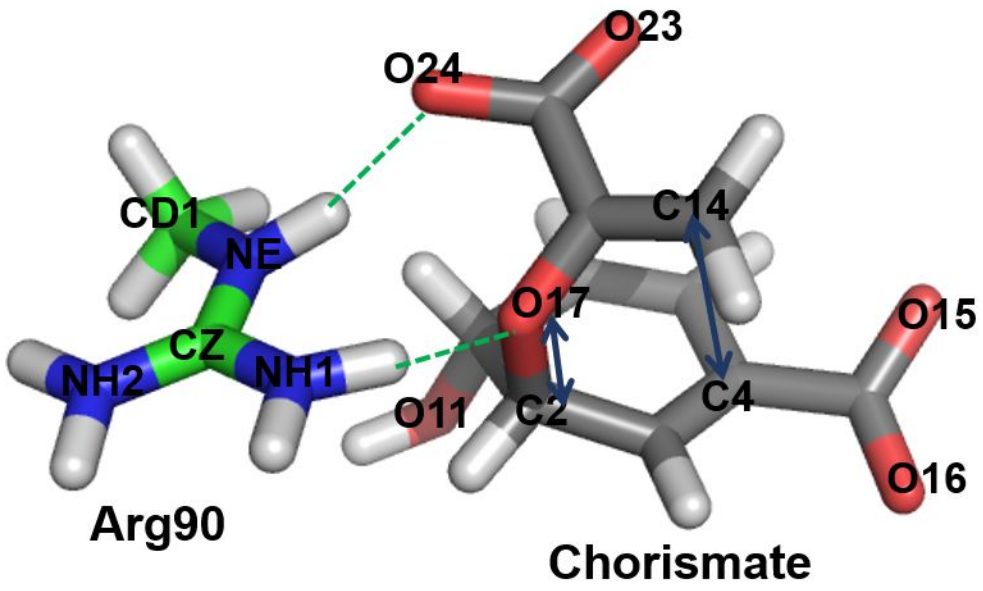



**Figure 1.** Atoms included in the QM region in BsCM, indicating atom names used in the text. The (blue) arrows indicate the distances involved in the reaction coordinate. The (green) dotted lines indicate hydrogen bonds formed between Arg90 and chorismate. In the initial simulations only chorismate was treated by QM and so the interaction between chorismate and Arg90 was treated at the QM/MM level. Subsequently, the size of the QM region was increased to include the side chain of Arg90, so treating this interaction (which is important for transition state stabilization in the enzyme[51, 54, 55]) with Arg90 by QM.

To explore the effect of QM region size in QM/MM calculations, reaction profiles were generated in solution treating different numbers of water molecules by QM in the QM/MM optimizations, allowing optimization of the internal structure of the water molecules. Figure 2 shows the different QM regions used, which range from a single QM water molecule to 16 QM water molecules. Water molecules were chosen based on distance from chorismate and in some cases form hydrogen bonds with it. In BsCM, we used 6 different starting structures (taken at 20 ps intervals from the 120 ps QM/MM MD simulation at the SCC-DFTB/CHARMM22 level) to calculate 6 reaction pathways with chorismate alone in the QM region (24 atoms, charge

−2*e*) and 6 pathways with chorismate and Arg90 (37 atoms including 1 link atom, charge −1*e*) treated by QM (see Figure 1). Further calculations were also performed for the enzyme with a still larger QM region, see below.

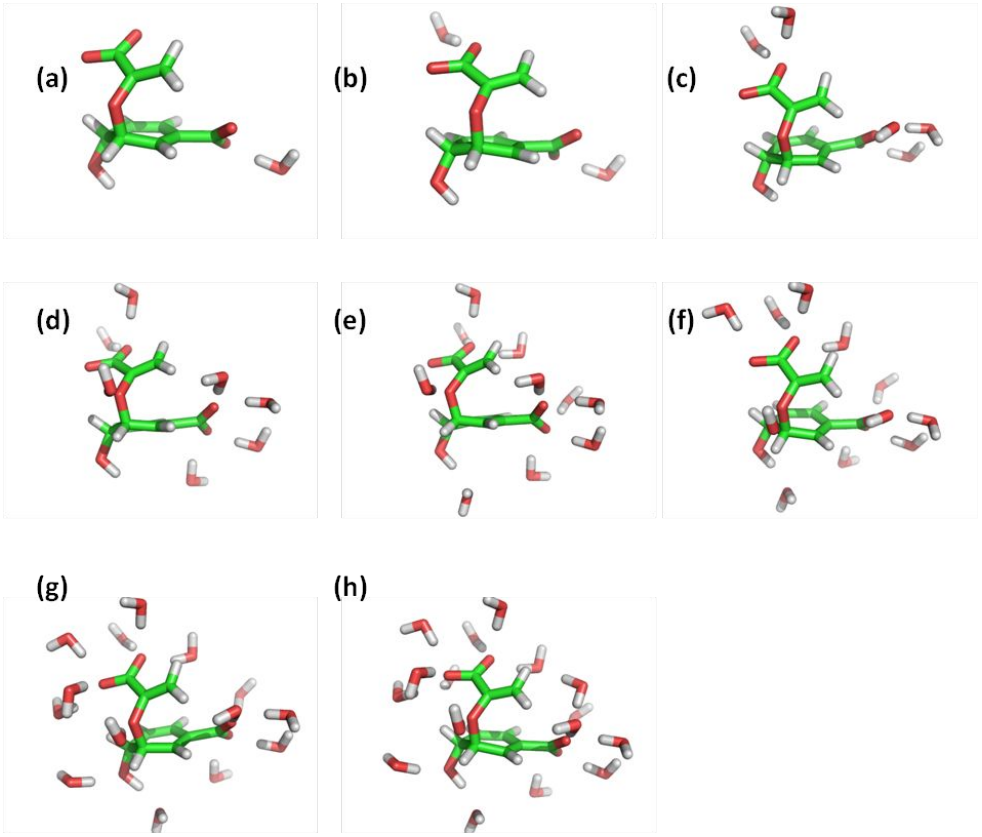



**Figure 2.** The different QM regions used in B3LYP QM/MM calculations of the conversion of chorismate to prephenate in water (MM atoms not shown). Increasing numbers of water molecules were treated by QM (in addition to chorismate) in each case: **(a)** water molecules within distance 1.64 Å (1 water molecule), **(b)** within 1.67 Å (2 water molecules), **(c)** within 1.71 Å (5 water molecules), **(d)** within 1.73 Å (7 water molecules), **(e)** within 1.74 Å (10 water molecules), **(f)** within 1.75 Å (11 water molecules), **(g)** within 1.77 Å (14 water molecules) and **(h)** within 1.80 Å (16 water molecules).

In typical (electrostatic embedding) QM/MM calculations (which use invariant point charge MM models), the partial atomic charges of the MM atoms do not change during the simulation. Effects such as polarization and charge transfer are not included for groups treated by MM. To gain insight into polarization, we calculated atomic charges in systems with different sizes of the QM region. In particular, we focused on how the partial atomic charges

vary along the reaction coordinate. For this, the simple Mulliken population analysis was used to calculate atomic charges for the QM atoms in the QM/MM calculations. We compared them with the partial atomic charges from the standard CHARMM27 forcefield used in the QM/MM calculations carried out to generate the reaction pathways. Mulliken population analysis has well-known limitations, including dependence on the basis set, and should not be taken as an indication of e.g. the electrostatic potential.[72] They can be useful, however, for indicating changes in electronic population during a reaction, or between different environments. In the analysis presented below, we calculate them at the level of theory used to generate the geometries and investigate changes along the pathway.

*Higher-level Corrections and Projector-based Embedding*

Single point energy calculations including the point charges of the enzyme or solution environment were performed using several different density functionals, and the Hartree-Fock method, for comparison. Functionals were chosen to test the effect of including exact exchange: M06-2X[73], BHLYP[74], M06[73], B3LYP[75], M06-L[76], and PBE[77] (in order of decreasing amounts of exact exchange). These methods are frequently used in QM and QM/MM studies of enzyme-catalysed reactions.[9, 30, 54, 78-81] All single point energy and embedding calculations were performed with the Molpro 2015.1 electronic structure program.[82] In the projector-based embedding calculations of the reaction in solution, the 24 atoms of chorismate were defined as sub-system A to be treated with the high-level WF method, while sub-system B encompassed an increasing number of water molecules (treated by DFT). Thus, in this case, where we refer to increasing numbers of water molecules, we are treating them only at the lower level (DFT) quantum method. These projector-based embedded calculations are further polarized by the MM potential of the TIP3P molecules in the water sphere. The calculations were carried out either using the cc-pVDZ basis or a combination of cc-pVTZ/aug-cc-pVTZ[83-85], in which the

diffuse functions were only present on oxygen atoms. For example, in the calculation for chorismate as sub-system A and seven water molecules in sub-system B, the cc-pVTZ/aug-cc-pVTZ basis reduced the number of molecular orbitals treated by the QM method from 94 to 59 and due to the basis truncation, the number of virtual orbitals was reduced from 1206 to 843 functions. The MP2[86], SCS-MP2[87], CCSD[88, 89] and CCSD(T)[90] correlated *ab initio* WF methods were used for the treatment of sub-system A, with B3LYP as the lower level DFT method used for sub-system B. Because of the close agreement (less than 1 kcal/mol difference) between CCSD(T) and SCS-MP2 for this reaction, we only performed the calculations at the SCS-MP2 level for the triple zeta basis.

This procedure was repeated for the BsCM model: again, chorismate was sub-system A treated by the high-level WF method, while an increasing number of arginine residues were added to sub-system B and treated at the DFT level (see Figure 3). These calculations were performed at the aug-cc-pVDZ level with SCS-MP2 for the high-level WF method. As the aim of these calculations was to assess the impact of QM region size on the calculated barrier, rather than to calculate the barrier with the greatest accuracy, the positions of the arginine residues in the 2ARG and 3ARG models were taken from QM/MM calculations with only Arg90 treated by QM, without any further geometry optimization, and a single point QM/MM calculation was performed to calculate the MM energy of the system. We note that the MM representation may not give correct treatment of the energetics for the changes in the internal structure of the arginine sidechain;[30]however, the internal geometry of arginine does not change much during the reaction. See the geometry of Arg90 in chorismate/TS/prephenate in a representative pathway where Arg90 was treated by MM and where it was included in the QM region in Figure S2. Note also that the charge of the QM region is changed by the inclusion of different numbers of arginine sidechains.

| | QM Arginines | Atoms in Sub-system A | Atoms in Sub-system B |
|---|---|---|---|
| 0ARG | | 24 | 0 |
| 1ARG | Arg90 | 24 | 13 |
| 2ARG | Arg90 + Arg7 | 24 | 26 |
| 3ARG | Arg90 + Arg7 + Arg63 | 24 | 39 |

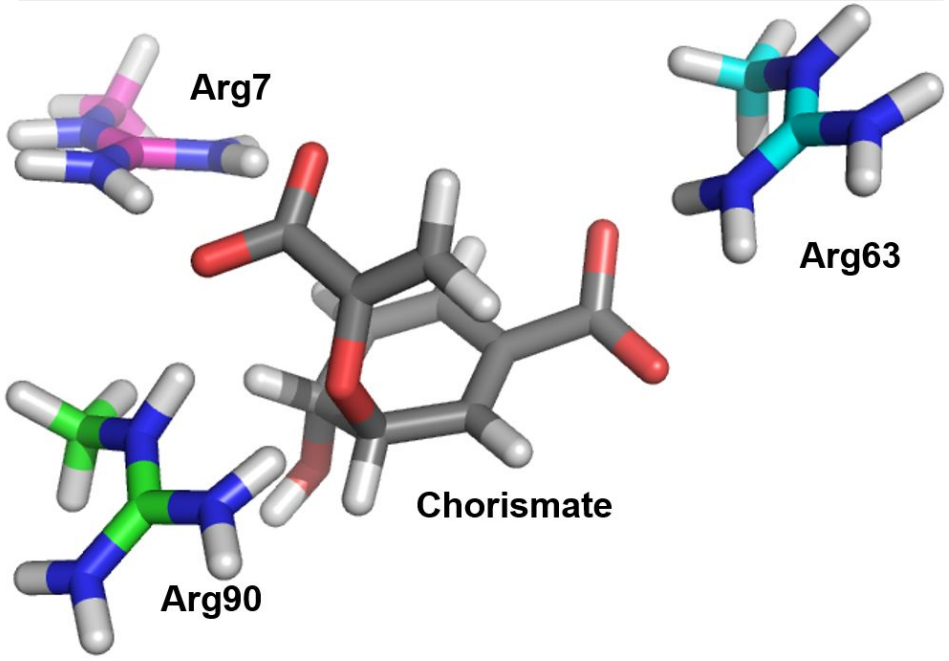



**Figure 3.** The different QM regions used in projector-based embedding calculations of the reaction in the BsCM enzyme. In all calculations, sub-system A contains only chorismate (grey carbons). One, two or three arginine residues are included in sub-system B and treated by DFT.

## RESULTS AND DISCUSSION

### *The rearrangement of chorismate to prephenate in solution*

The energetics of the (uncatalyzed) conversion of chorismate to prephenate in solution calculated at the B3LYP/6-31G(d)/CHARMM27 level using different QM region sizes are given in Table 1 (potential energy profiles are shown in Figure S3). The potential energy barrier is 19.6 kcal/mol when only the 24 atoms of chorismate are treated by B3LYP and all the water

by MM (CHARMM-type TIP3P). Increasing the size of the QM region from 24 to 72 atoms results in at most a ~ 2 kcal/mol difference in the potential energy barrier. There is very little variation in the calculated reaction energy as the size of QM region changes, with a difference of 0.6 kcal/mol between the smallest and largest QM regions. The change in energy and barrier is not linear with increasing number of QM atoms; changing the size of the QM region has very little effect (< 1 kcal/mol) on the calculated energy barrier until at least 10 water molecules are included in the QM region (cluster (e) in Figure 2) for which the barrier is reduced to 17.8 kcal/mol. Increasing the QM region size further to 14 QM water molecules reduces the barrier to 17.0 kcal/mol, but the barrier increases slightly again when further water molecules are added. There is only a weak correlation between the barrier height and reaction energy for these pathways ($R^2 = 0.40$). Overall, this shows that the interactions of water molecules are treated well by the QM/MM method, because the results are similar whether the water molecules are treated by QM or by MM.

The (approximate) TS occurs at a value of r = −0.5 Å on the reaction coordinate, with average C2 – O17 and C4 – C14 distances of 2.03 (±0.05) and 2.54 (±0.05) Å, respectively. The carboxylate to carboxylate distance is 4.98 (±0.04) Å at the TS. This is in excellent agreement with the data presented for the pathways reported by Claeyssens *et al*.[54] The structures of the QM region at the TS for the different models are shown in Figure S4, where it can be seen that there is very little difference in the geometry of the TS in for all sizes of QM region.

Table 2 shows the number of hydrogen bonds between the oxygen atoms of chorismate/TS/prephenate and the water molecules treated by QM in the different clusters. The water molecules in the clusters shown in Figure 2 were chosen based on their distance from chorismate, rather than any subjective choice or intuition about their role in stabilization. In the smaller QM clusters (< 7 QM water molecules), only the carboxylate oxygens (O15 & O23)

form hydrogen bonds with QM water molecules. It is not until cluster (e) that the hydroxyl group (O11) and ether oxygen (O17) are able to form hydrogen bonds with QM water molecules, which results in small decrease in barrier height.

**Table 1.** Potential energy barriers ($\Delta^{\ddagger}V$) for the conversion of chorismate to prephenate in water calculated at the B3LYP/6-31G(d)/CHARMM27 level of QM/MM theory with chorismate/TS/prephenate and an increasing number of water molecules treated by QM (see Figure 2 for the QM configurations).

| No. of water molecules | $\Delta^{\ddagger}V$ / kcal/mol | $\Delta_r V$ / kcal/mol |
|---|---|---|
| 0 | 19.6 | −14.1 |
| 1 | 20.0 | −13.9 |
| 2 | 19.5 | −12.9 |
| 5 | 19.6 | −12.7 |
| 7 | 19.0 | −13.3 |
| 10 | 17.8 | −14.4 |
| 11 | 17.4 | −13.9 |
| 14 | 17.0 | −14.4 |
| 16 | 17.7 | −14.7 |

In the MM forcefield all water oxygen atoms have a partial atomic charge of −0.834e and the hydrogen atoms have a partial atomic charge of 0.417e (TIP3P water model).[62, 65] The Mulliken charges of the oxygen atoms of the QM water molecules are generally more negative by up to 0.175e compared to the MM partial charge and the Mulliken charges of the hydrogen atoms are not identical [e.g. 1 QM water molecule: O = −0.987e, H1 = 0.428e and H2 = 0.445e, see Table S2]. Mulliken charge analysis indicates some transfer of negative charge from the chorismate/TS/prephenate to QM water molecules nearby. Figure S5 shows the sum of the partial atomic charges for the water molecules in the different clusters at the TS. The majority of the water molecules have a slightly negative overall charge, except the lime green colored

water molecule in clusters (f-h) which has a charge of 0.09e. Including the water molecules in the QM region allows the negative charge of the chorismate/TS/prephenate to be reduced by up to −0.729e (at the TS in the largest cluster), but cannot provide any positive charge to help stabilize the *TS*.

**Table 2.** The number of hydrogen bonds between the oxygen atoms of chorismate and water molecules treated by QM in the B3LYP/6-31G(d)/CHARMM27 reactant structures for each of the water clusters shown in Figure 2 (see Figure 1 for atom names). A hydrogen bond is defined as a hydrogen to acceptor distance < 2.5 Å and a donor–hydrogen–acceptor angle > 90°.[49]

| | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** |
|---|---|---|---|---|---|---|---|---|
| **O11** | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 2 |
| **O15** | 1 | 1 | 1 | 2 | 3 | 3 | 2 | 3 |
| **O16** | 0 | 0 | 2 | 2 | 2 | 2 | 3 | 3 |
| **O17** | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 2 |
| **O23** | 0 | 1 | 2 | 2 | 3 | 3 | 3 | 3 |
| **O24** | 0 | 0 | 0 | 0 | 0 | 1 | 3 | 3 |

*Higher level electronic structure with projector-based embedding*

So far in this work we have examined the reaction using the B3LYP density functional, which is known to give reasonable results for many reactions. Here, the potential energy barriers predicted by B3LYP/6-31G(d)/CHARMM27 method (17.0 to 20.0 kcal/mol) for the uncatalyzed reaction in solution are close to the experimental value of $\Delta^{\ddagger}H$ = 20.7 kcal/mol.[47] Thermal and ZPE contributions to the barrier are around –1.6 kcal/mol in BsCM[12]: adding these contributions to the calculated potential energy barriers $\Delta^{\ddagger}V$ gives estimates of the enthalpy barrier $\Delta^{\ddagger}H$ in the range 15.4 to 18.4 kcal/mol, slightly less than the experimental value. It has been noted previously that B3LYP underestimates the barrier for the chorismate to prephenate reaction.[12, 53, 54]

DFT is the workhorse of computational chemistry, and many modern functionals generally provide a good balance of accuracy and speed. However, DFT calculations (in contrast to *ab initio* methods) are not systematically improvable, and suffer from some well-known limitations. Different functionals can give very different results for the same reaction[32, 33], which can lead to qualitatively incorrect conclusions. Where possible, a standard approach to finding the best density functional for a particular application is to compare to correlated *ab initio* calculations on relevant small models or to draw from literature about similar chemical reactions.

Figure 4 shows the energy profiles for the uncatalyzed reaction in solution obtained from single point energy calculations with different density functionals (and HF) on the geometries obtained by optimization at the B3LYP/6-31G(d)/CHARMM27 level with chorismate and 16 water molecules as the QM region (72 QM atoms in total). As expected, Hartree-Fock calculations significantly overestimate the barrier, giving a barrier of 37.7 kcal/mol similar to that found in previous studies, and thus we do not discuss the HF results further here.[12, 51] The significant spread in the calculated barrier height (around 14 kcal/mol) with different density functionals shows the sensitivity of the barrier to the DFT functional. It also highlights how the proportion of exact exchange seems to be the most important factor. The two functionals with the largest amounts of exact exchange (M06-2X (54%) and BH&HLYP (50%)), give barriers of 24.6 and 26.0 kcal/mol, respectively. The next cluster of barrier heights is from the functionals with smaller exchange contributions: M06 (27%) and B3LYP (20%), with barriers of 19.1 and 17.0 kcal/mol, respectively. Finally, the exchange-free PBE and M06-L functionals give barriers of 11.8 and 12.3 kcal/mol. All the DFT methods find the transition state at a reaction coordinate value of −0.3 Å, apart from M06-L, for which it is located at a value of −0.2 Å on the reaction coordinate. There is less variation in the reaction energies predicted by the different DFT methods, with all the functionals giving a reaction

energy between –15 and –17 kcal/mol. HF gives a lower reaction energy of –21.4 kcal/mol. The product minimum is at a reaction coordinate value of $r$ = –1.4 Å for DFT and –1.6 Å for HF.

These results show that a spread of ~ 14 kcal/mol in barrier height is obtained depending on the choice of density functional. The question then remains as to which is the most reliable result? This question can be answered with projector-based embedding calculations, which can utilize almost any WF based method to obtain results to an arbitrary accuracy with reasonable cost. The major drawback of correlated *ab initio* methods is their computational cost, as most common methods have a formal scaling of at least two or three orders of magnitude greater then DFT. Various approaches have been developed to reduce the cost of WF methods, such as local correlation methods,[91-94] or highly efficient parallel implementations[14, 95, 96]. Local approximations applied in correlated calculations such as LCCSD or LMP2 exploit the short-range nature of dynamic correlation through the localization of molecular orbitals. They can give excellent results for reactions in large systems (e.g. in a QM/MM framework), but are not necessarily straightforward to apply (e.g. the definition of appropriate domains, which remain consistent cross the reaction coordinate, is a challenge particularly for non-expert users). Projector-based embedding is an attractive approach as it allows any WF method that uses an arbitrary core Hamiltonian to be embedded rigorously in a DFT environment.[32-34, 36]. This approach potentially allows for highly accurate correlated *ab initio* results for reactions in large systems, in a framework that is relatively straightforward to apply. Here we use this projector-based embedding approach in QM/MM calculations to achieve the accuracy of high-level correlated *ab initio* methods. For embedding calculations, the (DFT) QM region is partitioned further and a region to be treated by an *ab initio* method is chosen; both are polarized by the MM environment in a QM/MM calculation. Although the application in this work concerns non-covalent interactions it is entirely valid to make this 'cut' across covalent bonds in this

technique, as demonstrated in recent applications to citrate synthase.[32, 33] It has been shown that by using this approach, it is possible to largely eliminate the variability (and associated uncertainty) of DFT. At the same time, the effects of the wider environment can be taken into account within the QM/MM model. The projector-based embedding approach is significantly faster than correlated *ab initio* calculations including the whole QM region at the high-level because of the reduction of the number of electrons being excited and the number of virtual orbitals through basis truncation.[34] Indeed, for many typical QM/MM calculations, the CPU, memory and/or disk requirements make correlated *ab initio* calculations impossible. Currently, it is not possible to perform geometry optimization with projector-based embedding methods, due to the lack of gradients, so practical application requires the generation of structures with a lower level (e.g. DFT) method.

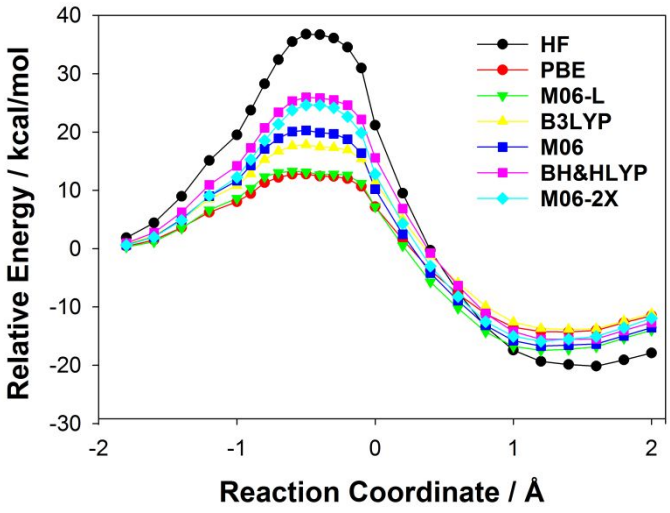



**Figure 4.** Potential energy profiles for the rearrangement of chorismate to prephenate in solution using different QM methods. The profiles are based on single point energy calculations with the designated QM method with the cc-pVDZ basis (using B3LYP/6-31G(d)/CHARMM27 geometries for the 72 QM atom pathways). All energies are relative to the reactant ($r = -1.8$ Å).

The flexibility of the projector-based embedding approach makes it possible to use a wide range of correlated *ab initio* methods for sub-system A. Figure 5 shows results of embedding different WF methods in B3LYP for the model of the reaction in solution, including

7 water molecules in sub-system B. The CCSD(T)-in-B3LYP/cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 barrier of 20.0 kcal/mol is taken as the benchmark here. Standard all-electron B3LYP/cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 gives a reaction barrier that is a little too low in comparison (17.8 kcal/mol). Also, the B3LYP potential energy surface is very flat around the transition state (reaction coordinate values −0.6 to 0.0 Å), while all the correlated *ab initio* methods show a more defined transition state with a sharper barrier and more curved surface. This emphasizes the need to treat with caution predictions of potential energy surface curvature from DFT calculations (e.g. for calculation of kinetic isotope effects[97, 98]). The barrier height at the SCS-MP2-in-B3LYP/cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 (19.8 kcal/mol) level is very similar to the CCSD(T) value (only 0.2 kcal/mol lower). The canonical MP2 barrier is 5.5 kcal/mol lower than CCSD(T). As found previously, SCS-MP2 compares more favorably than canonical MP2 to CCSD(T).[9, 11, 13] There is a significant difference in the energies predicted by the CCSD and CCSD(T) methods (~ 6 kcal/mol), showing that the inclusion of triple excitations is important. For the reaction energy, all the correlated *ab initio* methods give reaction energies within 1.5 kcal/mol of the CCSD(T) results. Notably, B3LYP calculates a reaction energy 6 kcal/mol higher (less exothermic) than any of the correlated *ab initio* methods. We also tested the effects of the size of the basis, which can be a critical factor in the accuracy of correlated methods, which may converge slowly with respect to the completeness of the basis. Figure S6 shows the barrier from the SCS-MP2-in-B3LYP/cc-pVTZ/aug-cc-pVTZ//B3LYP/6-31G(d)/CHARMM27 calculations for the 7 QM water case. We found that the barrier is lower by 1 kcal/mol with the larger basis and the reaction energy drops by 2.6 kcal/mol. From this point, we focus on using SCS-MP2 for our high-level analysis because it is in excellent agreement of 0.2 kcal/mol for the barrier with the CCSD(T) results and its similar curvature for the profile.

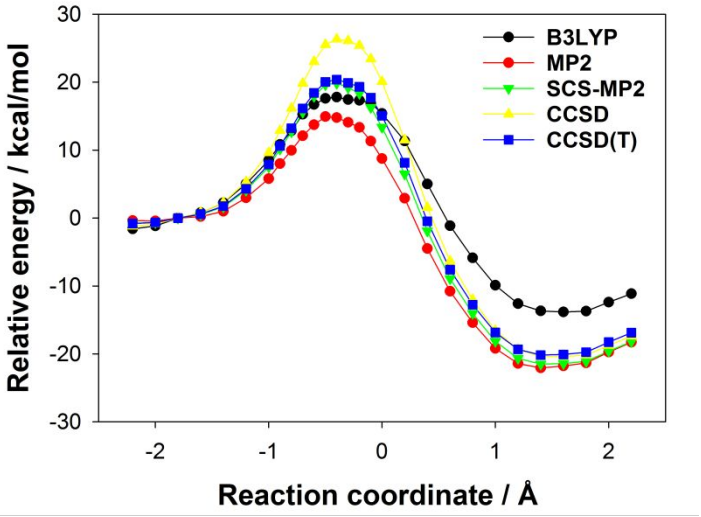



**Figure 5.** QM/MM potential energy profiles for the Claisen rearrangement of chorismate to prephenate in solution calculated using projector-based embedding of the *ab initio* method with the cc-pVDZ basis set (treating chorismate/TS/prephenate) in B3LYP for the QM region. For comparison the canonical B3LYP QM/MM result is also shown. The geometries are taken from B3LYP/6-31G(d)/CHARMM27 pathway calculations. In this system, chorismate and 7 water molecules are treated by QM (45 QM atoms in total). All energies are relative to the reactant ($r = -1.8$ Å).

The size of the quantum region size necessary for accurate QM/MM modelling of reactivity in condensed phase systems (such as enzymes and aqueous solution) is a topic of active discussion. As outlined in the introduction, some researchers have suggested that quite large QM regions are necessary to converge various chemical properties, such as reaction barriers.[19-23] Projector-based embedding in a QM/MM framework provides an accurate multiscale approach for calculating reaction energetics. We examined the effect of varying the number of QM waters in projector embedding QM/MM calculations of the uncatalysed reaction in aqueous solution, with SCS-MP2/cc-pVDZ for the high-level subsystem A (chorismate/TS/prephenate) (Figure 6). Note that the QM region has a charge of –2, so effects of charge transfer and polarization will be involved. The 0 QM water system (i.e. only the 24 atoms of chorismate/TS/prephenate as the QM region) gives a reaction barrier of 21.6 kcal/mol, whereas the model with 16 QM water molecules gives a barrier of 20.2 kcal/mol. Clearly the barrier is not very sensitive to the number of water

molecules treated by QM in this system. This shows that the QM/MM treatment of interactions with water molecules is a good approximation here; it also indicates that polarization and charge transfer do not change significantly during the reaction. The largest difference in reaction energetics found among all these different QM/MM systems is in the reaction energy between the 7 and 16 QM water cases, with the 7 QM water system being higher by 1.5 kcal/mol. The size of the QM region in the QM/MM calculations here has little effect on the barrier height in projector-based embedding QM/MM calculations (as also found in standard DFT/MM calculations above). The effect of changing the QM system size on the reaction energetics are relatively small, and certainly much smaller than the differences found when different density functionals are used (e.g. compare Figure 4).

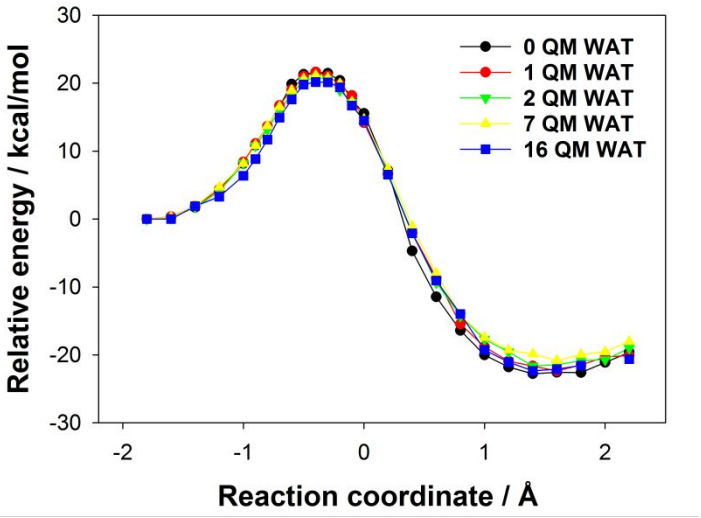



**Figure 6.** Potential energy profiles for the rearrangement of chorismate to prephenate in solution calculated using projector-based embedding of the SCS-MP2 method in B3LYP with different numbers of DFT water molecules in sub-system B, using the cc-pVDZ basis set. The geometries are taken from B3LYP/6-31G(d)/CHARMM27 pathway calculations treating chorismate and the stated number of water molecules by QM. All energies are relative to the reactant ($r = -1.8$ Å).

*Reaction in the enzyme*

Figure S7 shows the B3LYP/6-31G(d)/CHARMM27 potential energy profiles for the conversion of chorismate to prephenate in BsCM, generated with chorismate/TS/prephenate as the QM region or also including side chain of Arg90 in the QM region (shown in Figure S8). Arg90 is important for TS stabilization in the enzyme.[50, 54-57, 99] Six different energy profiles were generated for each system, from six snapshots taken from QM/MM MD simulations as described above. The potential energy profiles generated with these different sized QM regions have similar features: the reactant, TS and product located at the same value of the reaction coordinate. The TS is located at an average value of $r$ = − 0.5 Å in the pathways for both QM region sizes, in agreement with the location of the TS found by Claeyssens *et al*.[54] and the solution profiles presented here.

**Table 3.** A comparison of potential energy barriers ($\Delta^{\ddagger}V$) and reaction energies ($\Delta_r V$) calculated at the B3LYP/6-31G/CHARMM27 level of theory with chorismate/TS/prephenate (only) as the QM region; and with Arg90 also included in the QM region. The 6 pathways for each QM region size were generated from snapshots taken at 20 ps intervals from a 120 ps QM/MM MD simulation at the SCC-DFTB/CHARMM22 level. All energies are given in kcal/mol and the average value is also reported with the standard deviation given in parentheses.

| | **Chorismate QM** | | **Chorismate + Arg90 QM** | |
|---|---|---|---|---|
| **Path** | $\Delta^{\ddagger}V$ | $\Delta_r V$ | $\Delta^{\ddagger}V$ | $\Delta_r V$ |
| 1 | 11.9 | −18.2 | 11.2 | −18.1 |
| 2 | 12.4 | −18.3 | 12.1 | −16.8 |
| 3 | 12.8 | −17.6 | 11.8 | −17.1 |
| 4 | 11.5 | −17.2 | 11.3 | −16.5 |
| 5 | 13.5 | −17.5 | 12.5 | −14.8 |
| 6 | 9.9 | −20.7 | 9.2 | −19.7 |

| | | | | |
|---|---|---|---|---|
| **Average** | 12.0 (±1.3) | −18.3 (±1.3) | 11.4 (±1.2) | −17.1 (±1.7) |

The calculated energetics for this system are given in Table 3. The average potential barrier height when chorismate/TS/prephenate (only) is treated by QM is 12.0 (±1.3) kcal/mol, and slightly lower 11.4 (±1.2) kcal/mol when Arg90 is also included in the QM region (at the B3LYP/6-31G(d)/CHARMM27 level). The average reaction energy is –18.3 (±1.3) kcal/mol for the profiles where only chorismate/TS/prephenate is treated by QM and –17.1 (±1.7) kcal/mol for the larger QM region. It can be seen that the uncertainty due to conformational sampling is comparable to (and in fact somewhat larger than) the difference between the two different QM regions. There is some correlation between the barrier height and reaction energy: the pathways with the lowest barriers also have the most exothermic reaction energies [correlation coefficient between barrier and reaction energy $R^2 = 0.62$ chorismate/TS/prephenate only QM paths and $R^2 = 0.81$ for chorismate/TS/prephenate + Arg90 QM paths].

There is a small spread in the predicted barrier heights for the different starting structures, indicated by the ~1.3 kcal/mol standard deviation. Calculation of multiple profiles/barriers from multiple structures from MD simulations is a good way to test for effects of conformational variability, and potentially significance of results; comparison of multiple barriers can also e.g. identify structural determinants of reactivity.[100] The number of snapshots (e.g. starting structures from MD) necessary to achieve acceptable convergence in QM/MM potential energy calculations of reaction barriers is an important consideration, e.g. in making reliable predictions of regio-,

stereo- and chemoselectivity.[101-104] Ryde has discussed the number of snapshots needed to give a reliable estimate of the energy in QM/MM calculations of this type, and suggested that a value of 4 kJ/mol for the standard deviation (σ) is a good convergence criterion.[105] The 6 pathways initiated from snapshots from a QM/MM MD trajectory used here (Table 3) result in σ values of 1.3 kcal/mol (~5.4 kJ/mol) for the chorismate/TS/prephenate only QM paths and 1.2 kcal/mol (~5.0 kJ/mol) for the paths generated with chorismate/TS/prephenate and Arg90 treated by QM. Even for a simple reaction without any large-scale changes in the protein environment, a range of barriers will be calculated when different conformations are considered. Sampling protein structures e.g. through QM/MM or MM MD simulations is more computationally feasible when smaller QM regions are used. It should be noted that in some cases (e.g. where solvation changes significantly along the reaction coordinate), the calculations of potential energy profiles (e.g. by so-called adiabatic mapping) will not give useful or representative energy barriers or reaction energies. In such cases, sampling along the reaction coordinate is required. Free energy barriers (rather than potential energy barriers) can be calculated e.g. by MD simulations with enhanced sampling techniques (such as umbrella sampling or steered MD) and directly compared with experimental values.[102, 106-108] Such MD simulations typically cannot be carried out with high-level QM/MM methods (and certainly not with projector-based embedding QM/MM), but can give estimates of activation entropies and e.g. quantum tunneling corrections that can be combined with high-level QM/MM potential energy barriers.[12, 13, 97]

Modelling the interaction between chorismate/TS/prephenate and the side chain of Arg90 at the QM level, rather than by QM/MM, reduces the average barrier height by only 0.8 kcal/mol and reduces the exothermicity of the reaction by a similar amount (0.9 kcal/mol). Arg90 forms two hydrogen bonds to chorismate, which are maintained throughout the reaction: NH1 (HH11) – O17 and NE1 (HE1) – O24 (see Figure 1). The distances involved in these hydrogen bonds at the reactant, TS and product averaged over the 6 profiles calculated for each QM region size are shown in Table S3. The hydrogen bond between NH1 and O17 has an average hydrogen to acceptor distance of *d*(HH11 – O17) = 1.76 (±0.03) Å when Arg90 treated by MM and *d*(HH11 – O17) = 1.90 (±0.03) Å when Arg90 is treated by QM in the reactant. This hydrogen bond gets shorter at the TS, helping to stabilize the developing negative charge as the ether bond breaks[54, 109] [*d*(HH11 – O17) = 1.70 (±0.03) Å when Arg90 treated by MM; *d*(HH11 – O17) = 1.79 (±0.04) when Arg90 is treated by QM at the TS]. This hydrogen bond then lengthens in the product, returning to similar values to those observed in the reactant structures [*d*(HH11 – O17) = 1.75 (±0.03) Å when Arg90 treated by MM; *d*(HH11 – O17) = 1.89 (±0.04) Å when Arg90 is treated by QM in the product]. The hydrogen bond between NE1 and O24 lengthens slightly as the reaction progresses, increasing from an average hydrogen to acceptor distance of *d*(HE1 – O24) = 1.87 (±0.08) Å when Arg90 treated by MM and *d*(HE1 – O24) = 1.90 (±0.07) Å when Arg90 is treated by QM in the reactant, to *d*(HE1 – O24) = 1.97 (±0.12) Å when Arg90 treated by MM; *d*(HE1 – O24) = 2.03 (±0.09) Å when Arg90 is treated by QM in the product. These hydrogen bonds with Arg90 are slightly shorter (~0.1 Å) when this residue is treated by MM rather than QM methods. It is apparent that the QM/MM

treatment of the interaction is not exactly the same as the QM treatment, as expected, but the change in this interaction during the reaction is modelled well by QM/MM (because it is primarily an electrostatic interaction[109]). Biomolecular hydrogen bond interactions can be modelled well by QM/MM methods.[15, 16]

Table 4 shows a comparison of the partial atomic charges of the side chain of an arginine residue in the CHARMM27 forcefield[71] (as used in the Chorismate/TS/prephenate only QM calculations) and the partial atomic charges of Arg90 at the reactant (R), transition state (TS) and product (P), from Mulliken population analysis,[72] from QM/MM calculations at the B3LYP/6-31G(d)/CHARMM27 level of theory where Arg90 is included in the QM region. Mulliken population analysis shows that the partial atomic charges of the arginine atoms show very little variation along the reaction path, even for those atoms involved in hydrogen bonds. For example, the Mulliken charge of HH11 is 0.35*e* at R and in the TS and 0.36*e* in P, despite this hydrogen bond shortening by 0.11 Å (as described above). The partial atomic charges of the nitrogen atoms are the same when treated by QM or in the MM representation at −0.8*e*. One difference between the MM and QM partial atomic charges is that both hydrogens of the NH2 groups are of the same charge (0.46*e*) in the MM forcefield, but they have significantly different charges in the QM population analysis e.g. HH11 = 0.35*e* and HH12 = 0.47*e*. On the whole, the Mulliken charges of most Arg90 atoms are similar to those in the MM forcefield, which is some indication of reasonable consistency between the MM and QM treatments of this residue (e.g. the partial atomic charge of NE is slightly less negative at the QM level at −0.62*e*, compared to −0.7*e* in the CHARMM27 forcefield; the partial atomic charge of HE is also less positive at the QM level (0.44*e* MM, 0.41*e* QM)). Table S4 shows the partial atomic charges of the oxygen atoms of chorismate/TS/prephenate from Mulliken population analysis with and without Arg90 in the QM region. Again, the differences are very small even for O17 and O24 that form hydrogen bonds with Arg90 (see Figure 1). The magnitude of the negative

partial charge of O17 is 0.05*e* smaller in R and 0.07*e* smaller in the TS and P. As noted above, the interaction with Arg90 (which contributes significantly to TS stabilization) is largely electrostatic[55-57, 109] and is adequately represented by the point charge model in the forcefield.

**Table 4.** Partial atomic charges used for the atoms of the side chain of Arg90 in the CHARMM27 forcefield[65] and the (average, over 6 profiles) partial atomic charges from Mulliken population analysis[72] in the B3LYP/6-31G(d)/CHARMM27 QM/MM calculations of the reactant (R), transition state (TS) and product (P). All charges are given in atomic units (*e*).

| ATOM NAME | CHARMM27 charge | Mulliken Population B3LYP/6-31G(d)/CHARMM27 | | |
|---|---|---|---|---|
| | | R | TS | P |
| CD | 0.2 | −0.33 | −0.33 | −0.34 |
| HD1 | 0.09 | 0.19 | 0.18 | 0.19 |
| HD2 | 0.09 | 0.22 | 0.22 | 0.23 |
| NE | −0.7 | −0.62 | −0.63 | −0.63 |
| HE | 0.44 | 0.41 | 0.41 | 0.40 |
| CZ | 0.64 | 0.82 | 0.81 | 0.82 |
| NH1 | −0.8 | −0.80 | −0.80 | −0.80 |

| | | | | |
|---|---|---|---|---|
| HH11 | 0.46 | 0.35 | 0.35 | 0.36 |
| HH12 | 0.46 | 0.47 | 0.47 | 0.47 |
| NH2 | −0.8 | −0.80 | −0.81 | −0.80 |
| HH21 | 0.46 | 0.41 | 0.42 | 0.42 |
| HH22 | 0.46 | 0.38 | 0.38 | 0.38 |

*High-level electronic structure QM/MM calculations with projector-based embedding*

Figure 7(a) shows the DFT/aug-cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 (standard DFT/MM, i.e. not projector-based embedding) energy profiles for a reaction pathway with chorismate/TS/prephenate and Arg90 treated by QM. As found above for QM/MM calculations on the (uncatalyzed) reaction in solution Figure 4, there is a ~13 kcal/mol spread in the potential energy barrier for this reaction calculated with different density functionals e.g. ($\Delta^{\ddagger}V$ = 7.30 kcal/mol with M06/aug-cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 and ($\Delta^{\ddagger}V$ = 20.0 kcal/mol with M06-2X/aug-cc-pVDZ//B3LYP/6-31G(d)/CHARMM27. The reaction energy varies from $\Delta_r V$ = −15.8 kcal/mol at the B3LYP/aug-cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 level to $\Delta_r V$ = −18.6 kcal/mol at the M06-L/aug-cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 level. The position of the TS on the reaction coordinate differs between methods, with the $r$ = −0.6 Å M06-L with the method, $r$ = −0.5 Å with the PBE and B3LYP methods, and $r$ = −0.4 Å with the BH&HLYP, M06 and M06-2X methods. These DFT/CHARMM27 results for the potential energy barrier for the enzymatic reaction are 5-6 kcal/mol lower than the equivalent values for the solution reaction, while the values for the reaction energy are similar.

We also applied projector-based embedding techniques to calculate the energetics of the reaction in BsCM. When projector-based embedding techniques are

applied to treat sub-system A with SCS-MP2 with sub-system B included by the same range of DFT functionals (Figure 7(b)), the spread in energy barriers predicted by the different density functionals is eliminated and there is only a 0.3 kcal/mol difference in the calculated barrier [$\Delta^{\ddagger}V$ = 15.3 kcal/mol at the SCS MP2-in-PBE/aug-cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 level and [$\Delta^{\ddagger}V$ = 15.6 kcal/mol at the SCS MP2-in-M06-L/aug-cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 level]. If thermal and ZPE contributions to the barrier of –1.6 kcal/mol are considered[12], then our estimate of $\Delta^{\ddagger}H$ = 13.7 – 14.0 kcal/mol, in good agreement with the experimental value of $\Delta^{\ddagger}H$ = 12.7±0.4 kcal/mol.[47] The difference in energy barrier between the enzyme (~14 kcal/mol) and solution (~18.4 kcal/mol) is 4.4 kcal/mol, somewhat lower than the expected $\Delta\Delta^{\ddagger}H$ = 8 kcal/mol.[47] As for the reaction in solution, the *ab initio* potential energy profile shows a more defined transition state located at the same value of the reaction coordinate with all methods ($r$ = −0.4 Å). There is also very little difference in the values for the reaction energy, with more exothermic values when projector-based embedding is included in the calculation, ranging from −20.6 to −21.4 kcal/mol. The reaction energy is similar for the reaction in both the enzyme and solution with projector-based embedding included in the energy calculations.

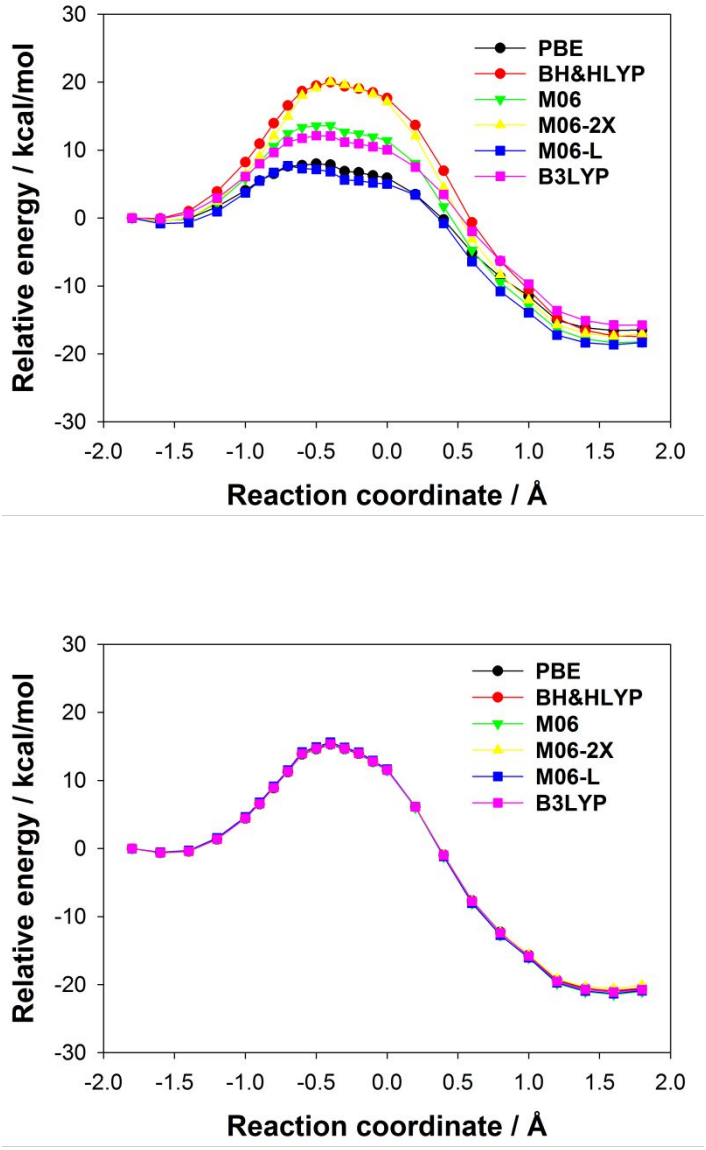
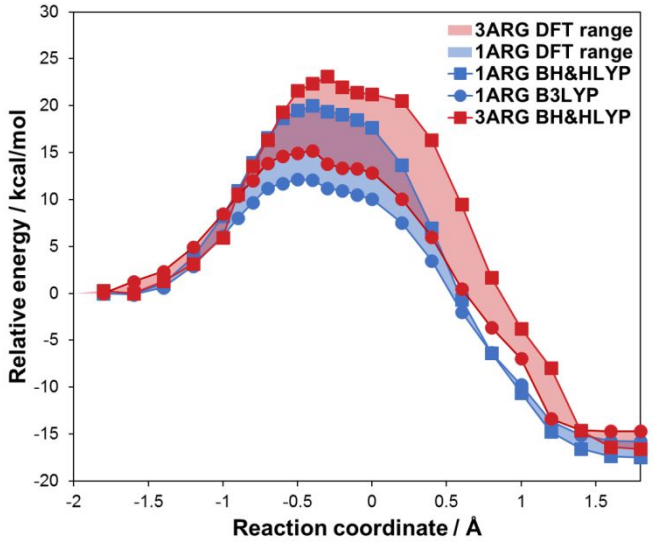



**Figure 7. (a)** Potential energy profiles for the rearrangement of chorismate to prephenate in BsCM at the DFT/aug-cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 level of theory, using the PBE, BH&HLYP, M06, M06-2X, M06-L and B3LYP functionals. **(b)** Projector-based embedding profiles at the SCS-MP2-in-DFT/aug-cc-pVDZ// B3LYP/6-31G(d)/CHARMM27 level of theory for the same range of DFT functionals. All energies are relative to the reactant ($r = -1.8$ Å).

To test the impact of QM region size on these QM/MM calculations we included 1 or 2 further arginine side chains in the QM region (without further QM/MM optimization of the profiles). Figure 8(a) shows the profiles obtained for the 1 QM arginine (1ARG: Arg90) and 3 QM arginine side chains (3ARG: Arg90, Arg7 +

Arg63) cases, respectively, calculated with B3LYP/aug-cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 and BH&HLYP/aug-cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 QM/MM methods. For clarity, only the 1ARG and 3ARG cases are discussed here. Profiles for all the QM region sizes at the B3LYP/aug-cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 are shown in Figure S9 and the corresponding SCS-MP2-in-B3LYP profiles are shown in Figure S10. For both the B3LYP and BH&HLYP methods increasing the size of the QM region by the addition of 2 further arginine residues (26 atoms, changing the charge on the QM region from –1*e* to +1*e*) increases the calculated barrier height by ~ 3 kcal/mol. To test the dependence of the energetics in these systems on the density functionals, Figure 8a shows comparable profiles with either 1 arginine or 3 arginine residues in the QM region calculated with BH&HLYP and B3LYP QM/MM methods. Figure 8a shows in blue the range of energies for the chorismate + Arg90 QM case (1ARG) case calculated with the BH&HLYP (squares) and B3LYP (circles) methods (i.e. the difference between the two functionals), also shown in Figure 7(a). The equivalent profiles for the 3 arginine QM system are shown in red in Figure 8a. Comparing these results shows that the barrier is about twice as sensitive to the functional (5.7 kcal/mol at $r = -0.2$ Å) as it is to changes in the size QM region (2.7 kcal/mol for B3LYP/MM). The shape of the potential energy profile shapes also strongly depends on the functional. Figure 8(b) shows the results of embedded QM/MM calculations (SCS-MP2-in-B3LYP/aug-cc-pVDZ//B3LYP/6-31G(d)/CHARMM27) for the same profiles. The SCS-MP2-in-B3LYP/aug-cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 barrier is 15.4 kcal/mol for the 1ARG model and 18.6 kcal/mol for the 3ARG case, a difference in energy of 3.2

kcal/mol (at $r$ = −0.4 Å on the reaction coordinate). This value of 18.6 kcal/mol (17.0 kcal/mol if thermal and ZPE contributions are included) is somewhat higher than the experimental value of $\Delta^{\ddagger}H$ = 12.7 (±0.4) kcal/mol and similar to the results for the uncatalyzed reaction. It should be emphasized that the QM region has not been optimized in this case; the larger QM region is used purely as a test of energies. It is known that the MM description of arginine is not correct, and can affect QM/MM energies[30] and reliable energetics would require QM/MM optimization.



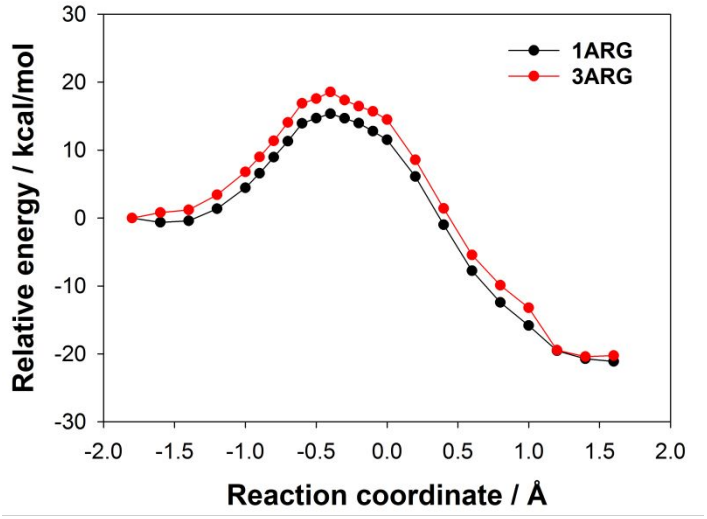



**Figure 8. (a)** Potential energy profiles for the rearrangement of chorismate to prephenate in BsCM for different QM region sizes calculated with commonly used functionals B3LYP and BH&HLYP (all at the DFT/aug-cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 level). All energies are relative to the reactant ($r = -1.8$ Å). The shaded area represents the difference in energy between the two functionals for that system. **(b)** SCS-MP2-in-B3LYP/aug-cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 profiles for the different QM region sizes. All energies are relative to the reactant ($r = -1.8$ Å).

## CONCLUSIONS

We have carried out QM/MM single point energy calculations using a range of density functionals on structures generated by QM/MM adiabatic mapping along a defined reaction coordinate to model the rearrangement of chorismate to prephenate in solution and in the BsCM enzyme. The profiles were initiated from structures of the TS from QM/MM umbrella sampling MD simulations at the SCC-DFTB/CHARMM22 level of theory. We have tested the sensitivity of the results to the choice of DFT functional, *ab initio* method and QM region size. The results highlight the significant differences in barriers and reaction energies given by different commonly used hybrid functionals (B3LYP and BHLYP) and meta hybrids (M06 and M06-2x): these various functionals predict barriers which differ by as much as 8.4 kcal/mol for the reaction in solution. We compared the barrier to reaction in solution calculated at the B3LYP

pVDZ//B3LYP/6-31G(d)/CHARMM27 level to the results of projector-based embedding of various *ab initio* methods. The CCSD(T)-in-B3LYP/cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 barrier of 20.0 kcal/mol is our benchmark for these calculations. Thermal and ZPE contributions (calculated at the B3LYP level) of –1.6 kcal/mol give an estimate of $\Delta^{\ddagger}H$ = 18.4 kcal/mol, close to the experimental value of $\Delta^{\ddagger}H$ = 20.7 kcal/mol.[47] The SCS MP2-in-B3LYP/cc-pVDZ//B3LYP/6-31G(d)/CHARMM27 barrier of 19.8 kcal/mol is in excellent agreement with the CCSD(T) value as has been shown previously. MP2 and CCSD methods predicted barriers ~5 kcal/mol lower than the CCSD(T) result. Consequently, all subsequent projector-based embedding calculations were carried out at the SCS MP2 level of theory.

Calculations on the reaction in the enzyme (BsCM), also showed large differences in barriers between different functionals (e.g. the difference in barrier the commonly used functionals B3LYP and BHLYP is ~ 8 kcal/mol (for the a chorismate/TS/prephenate + Arg90 QM system). This variation in barrier height with density functional is effectively removed by projector-based embedding: the differences between barriers with different functionals is only ~0.3 kcal/mol at the SCS-MP2-in-DFT/aug-cc-pVDZ/B3LYP/63-1G(d)/CHARMM27 level.

The shape of the potential energy surface was also different: the more accurate *ab initio* methods give a more curved profile and well-defined transition state, compared to the B3LYP/6-31G(d)/CHARMM27 profile, which is quite flat in the region of the transition state. The reaction energy at the B3LYP/ aug-cc-pVDZ/B3LYP/6-31G(d)/CHARMM27 level is also less exothermic by 6 kcal/mol compared to SCS MP2 or CCSD(T) calculations.

We investigated the sensitivity of the results to the size of QM region. In solution we modelled chorismate/TS/prephenate and increasing numbers of water molecules by QM. The variation of the barrier due to the size of the QM region was 3 kcal/mol in the B3LYP/6-

31G(d)/CHARMM27 profiles. In the SCS MP2-in-B3LYP/aug-cc-pVDZ/B3LYP/6-31G(d)/CHARMM27 calculations the barriers differ by ~1.5 kcal/mol. In the enzyme the average difference in energy barrier between the model where only chorismate/TS/prephenate is treated by QM and one with Arg90 included in the QM region is 0.8 kcal/mol. The stabilization provided by Arg90 is mostly electrostatic in nature[109] and the point charge representation of arginine in the forcefield provides a reasonably accurate description of this interaction, hence the small difference in energy. Larger QM models with additional arginine residues (Arg7 and Arg63) in the QM region were also investigated (without any optimization of the geometry). Increasing the number of arginine residues increased the barrier by ~3 kcal/mol at all levels of theory, but the reaction energy is much less sensitive to the size of the QM region. The difference in barrier heights predicted by common density functionals for any size of QM region is significantly larger than the change in barrier caused by increasing the number of arginine residues treated in the QM region. For the 3ARG model, the SCS MP2 in-B3LYP/aug-cc-pVDZ/B3LYP/6-31G(d)/CHARMM27 barrier is 18.3 kcal/mol, but this may be reduced by full optimization of the QM region.

Sampling of conformations is also important to draw reliable conclusions from QM/MM calculations on enzyme reactions. For six different structures of the enzyme taken from QM/MM MD simulations, the variation in energy was ~ 1.3 kcal/mol. This enzyme active site is relatively well-ordered and no significant conformational changes are required for reaction, but nevertheless, thermal fluctuations will give rise to small fluctuations in the barrier. The effect of conformational variation here is relatively small, and comparable the effect of increasing the QM region by 13 atoms to include Arg90 in the B3LYP/6-31G(d)/CHARMM27 pathway calculations. A practical consideration is that smaller QM regions allow more extensive sampling of

conformations (e.g. in MD simulations), which is also an important factor in the reliability of mechanistic insight that has been gained through many 'small' QM region QM/MM studies. The size of QM region should be balanced against the amount of sampling required to test mechanistic conclusions or predictions of selectivity.[104, 110]

There are a number of factors that should be considered and tested in any QM/MM investigation, including the choice of QM method, the size of the QM region and the amount of sampling required. Generally, in practical investigations of reactivity, relative barriers (rather than absolute barriers) are important,[111, 112] e.g. to distinguish between possible mechanisms, to identify reactive conformations[113, 114], predict the effects of mutation, reactivity of alternative substrates or analyse chemo-, stereo- or regioselectivity.[101-104] However, limitations of DFT may prevent reliable predictions even for calculations of relative barriers. Multiple applications have shown that, when applied sensibly and appropriately, QM/MM calculations can provide insight and can make useful predictions,[115-118] and e.g. identify catalytic interactions[119, 120]. For some applications, errors intrinsic to DFT (as currently applied), or uncertainty due to differences between functionals, may be the most significant limitation in making reliable predictions. For the systems studied here, the reaction barrier is significantly more sensitive to the functional than it is to the size of the QM region. It is useful to test the sensitivity of the results to the size of the QM region, but it should be remembered that the choice of the density functional may be more of a limiting factor in accuracy. Dispersion effects within the QM region may be important[121] and may need to be included in DFT calculations (e.g. via empirical corrections[122, 123]). Also QM/MM methods are intrinsically approximate (the approximations and limitations of the MM forcefield are probably now a limiting factor in accuracy) and simple convergence of properties with increasing size of the QM region cannot necessarily be expected: e.g., errors of electronic polarization and charge

transfer, not represented in typical MM forcefields, are often relatively small for small QM regions[15] but may increase with the size of the QM region and associated increase in size of the QM/MM boundary. Use of large QM regions (100s of atoms) may require somewhat different treatment of the boundary, e.g. frozen density methods[28, 29]. For reactions in chorismate mutase, there is no evidence of significant discrepancy between large-scale (up to 2000 atoms) pure QM results and QM/MM calculations.[48] QM/MM methods have also been shown to be able to predict effects of mutation, and analyse the origins of catalysis, in this enzyme. [38, 40, 42, 49-54, 57, 58]

Highly accurate (and systematically improvable/testable) calculations of potential energy surfaces and reaction barriers require correlated *ab initio* methods.[12, 124] Only with high-levels of *ab initio* theory (e.g. CCSD(T)) can barriers generally be predicted to close to chemical accuracy (1 kcal/mol) in electronic structure calculations. SCS-MP2 calculations can be close to CCSD(T) in accuracy, as found here and previously,[9-11, 13] and provide a good balance of accuracy and cost. Such methods can be applied in large systems using local approximations (e.g. LCCSD(T) etc.[12]), but these currently can require expert knowledge for useful calculations (e.g. in the definition of domains). Projector-based embedding[31] provides a more straightforward, easier to use approach for high-level QM and QM/MM calculations in chemistry and enzymology.[32, 33] Practically, structures generated with lower level methods (e.g. DFT/MM) must be used. A relatively small region encompassing the reaction is embedded rigorously in a large QM region treated by DFT; the effects of the bulk of the system can be included in a QM/MM framework. Barriers calculated by projector-based embedding are not sensitive to the functional used for embedding, allowing accurate calculations of reaction barriers and potentially other properties at reasonable computational cost.

## ACKNOWLEDGEMENTS

AJM thanks EPSRC (grant numbers EP/M022609/1, CCP-BioSim (ccpbiosim.ac.uk) and EP/M013219/1) and with KER thanks the BBSRC for funding (grant number BB/ M000354/1). SJB thanks EPSRC (with AJM) for grant EP/M022129/1, HECBioSim (hecbiosim.ac.uk) and the School of Chemistry, University of Bristol for funding. NL thanks the Thai Government for Higher Educational Strategic Scholarships for Frontier Research Network (SFR) scholarship. DS is grateful for support from the EPSRC Centre for Doctoral Training in Theory and Modelling in Chemical Sciences (Grant No. EP/L015722/1). This work was carried out using the computational facilities of the Advanced Computing Research Centre at the University of Bristol http://www.bris.ac.uk/acrc.

## SUPPORTING INFORMATION

Details of the nudged elastic band calculations to carried out to test the reaction coordinate are available in section 1.1 of the SI. Figures S1 to S10 and Tables S1-S3 are also included.

REFERENCES

1. Warshel, A., Multiscale Modeling of Biological Functions: From Enzymes to Molecular Machines (Nobel Lecture). *Angewandte Chemie-International Edition* **2014,** *53* (38), 10020-10031.

2. Karplus, M., Development of Multiscale Models for Complex Chemical Systems: From H+H-2 to Biomolecules (Nobel Lecture). *Angewandte Chemie-International Edition* **2014,** *53* (38), 9992-10005.

3. Levitt, M., Birth and Future of Multiscale Modeling for Macromolecular Systems (Nobel Lecture). *Angewandte Chemie-International Edition* **2014,** *53* (38), 10006-10018.

4. Amaro, R. E.; Mulholland, A. J., Multiscale methods in drug design bridge chemical and biological complexity in the search for cures. *Nature Reviews Chemistry* **2018,** *2* (4).

5. Piccini, G.; Alessio, M.; Sauer, J., Ab initio study of methanol and ethanol adsorption on BrOnsted sites in zeolite H-MFI. *Physical Chemistry Chemical Physics* **2018,** *20* (30), 19964-19970.

6. O'Malley, A. J.; Logsdail, A. J.; Sokol, A. A.; Catlow, C. R. A., Modelling metal centres, acid sites and reaction mechanisms in microporous catalysts. *Faraday Discussions* **2016,** *188*, 235-255.

7. Hohenberg, P.; Kohn, W., Inhomogeneous Electron Gas. *Physical Review B* **1964,** *136* (3B), 864-871.

8. Kohn, W.; Sham, L. J., Self-consistent equations including exchange and correlation effects. *Physical Review* **1965,** *140* (4A), 1133-1138.

9. Lawan, N.; Ranaghan, K. E.; Manby, F. R.; Mulholland, A. J., Comparison of DFT and ab initio QM/MM methods for modelling reaction in chorismate synthase. *Chemical Physics Letters* **2014,** *608*, 380-385.

10. van der Kamp, M. W.; Perruccio, F.; Mulholland, A. J., High-level QM/MM modelling predicts an arginine as the acid in the condensation reaction catalysed by citrate synthase. *Chemical Communications* **2008,** (16), 1874-1876.

11. van der Kamp, M. W.; Zurek, J.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J., Testing High-Level QM/MM Methods for Modeling Enzyme Reactions: Acetyl-CoA Deprotonation in Citrate Synthase. *Journal of Physical Chemistry B* **2010,** *114* (34), 11303-11314.

12. Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schuetz, M.; Thiel, S.; Thiel, W.; Werner, H.-J., High-accuracy computation of reaction barriers in enzymes. *Angewandte Chemie-International Edition* **2006,** *45* (41), 6856-6859.

13. Ranaghan, K. E.; Morris, W. G.; Masgrau, L.; Senthilkumar, K.; Johannissen, L. O.; Scrutton, N. S.; Harvey, J. N.; Manby, F. R.; Mulholland, A. J., Ab Initio QM/MM Modeling of the Rate-Limiting Proton Transfer Step in the Deamination of Tryptamine by Aromatic Amine Dehydrogenase. *Journal of Physical Chemistry B* **2017,** *121* (42), 9785-9798.

14. Bistoni, G.; Polyak, I.; Sparta, M.; Thiel, W.; Neese, F., Toward Accurate QM/MM Reaction Barriers with Large QM Regions Using Domain Based Pair

Natural Orbital Coupled Cluster Theory. *Journal of Chemical Theory and Computation* **2018,** *14* (7), 3524-3531.

15. Senthilkumar, K.; Mujika, J. I.; Ranaghan, K. E.; Manby, F. R.; Mulholland, A. J.; Harvey, J. N., Analysis of polarization in QM/MM modelling of biologically relevant hydrogen bonds. *Journal of the Royal Society Interface* **2008,** *5*, S207-S216.
16. Pentikainen, U.; Shaw, K. E.; Senthilkumar, K.; Woods, C. J.; Mulholland, A. J., Lennard-Jones Parameters for B3LYP/CHARMM27 QM/MM Modeling of Nucleic Acid Bases. *Journal of Chemical Theory and Computation* **2009,** *5* (2), 396-410.
17. Shaw, K. E.; Woods, C. J.; Mulholland, A. J., Compatibility of Quantum Chemical Methods and Empirical (MM) Water Models in Quantum Mechanics/Molecular Mechanics Liquid Water Simulations. *Journal of Physical Chemistry Letters* **2010,** *1* (1), 219-223.
18. Woods, C. J.; Shaw, K. E.; Mulholland, A. J., Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Simulations for Protein-Ligand Complexes: Free Energies of Binding of Water Molecules in Influenza Neuraminidase. *Journal of Physical Chemistry B* **2015,** *119* (3), 997-1001.
19. Kulik, H. J.; Zhang, J. Y.; Klinman, J. P.; Martinez, T. J., How Large Should the QM Region Be in QM/MM Calculations? The Case of Catechol O-Methyltransferase. *Journal of Physical Chemistry B* **2016,** *120* (44), 11381-11394.
20. Flaig, D.; Beer, M.; Ochsenfeld, C., Convergence of Electronic Structure with the Size of the QM Region: Example of QM/MM NMR Shieldings. *Journal of Chemical Theory and Computation* **2012,** *8* (7), 2260-2271.

21. Sumowski, C. V.; Ochsenfeld, C., A Convergence Study of QM/MM Isomerization Energies with the Selected Size of the QM Region for Peptidic Systems. *Journal of Physical Chemistry A* **2009,** *113* (43), 11734-11741.

22. Rossbach, S.; Ochsenfeld, C., Influence of Coupling and Embedding Schemes on QM Size Convergence in QM/MM Approaches for the Example of a Proton Transfer in DNA. *Journal of Chemical Theory and Computation* **2017,** *13* (3), 1102-1107.

23. Kulik, H. J., Large-scale QM/MM free energy simulations of enzyme catalysis reveal the influence of charge transfer. *Physical Chemistry Chemical Physics* **2018,** *20* (31), 20650-20660.

24. Jindal, G.; Warshel, A., Exploring the Dependence of QM/MM Calculations of Enzyme Catalysis on the Size of the QM Region. *Journal of Physical Chemistry B* **2016,** *120* (37), 9913-9921.

25. Karelina, M.; Kulik, H. J., Systematic Quantum Mechanical Region Determination in QM/MM Simulation. *Journal of Chemical Theory and Computation* **2017,** *13* (2), 563-576.

26. Das, S.; Nam, K.; Major, D. T., Rapid Convergence of Energy and Free Energy Profiles with Quantum Mechanical Size in Quantum Mechanical-Molecular Mechanical Simulations of Proton Transfer in DNA. *Journal of Chemical Theory and Computation* **2018,** *14* (3), 1695-1705.

27. Liao, R. Z.; Thiel, W., Convergence in the QM-Only and QM/MM Modeling of Enzymatic Reactions: A Case Study for Acetylene Hydratase. *Journal of Computational Chemistry* **2013,** *34* (27), 2389-2397.

28. Wesolowski, T. A.; Warshel, A., FROZEN DENSITY-FUNCTIONAL APPROACH FOR AB-INITIO CALCULATIONS OF SOLVATED MOLECULES. *Journal of Physical Chemistry* **1993,** *97* (30), 8050-8053.

29. Zech, A.; Ricardi, N.; Prager, S.; Dreuw, A.; Wesolowski, T. A., Benchmark of Excitation Energy Shifts from Frozen-Density Embedding Theory: Introduction of a Density-Overlap-Based Applicability Threshold. *Journal of Chemical Theory and Computation* **2018,** *14* (8), 4028-4040.

30. Kaiyawet, N.; Lonsdale, R.; Rungrotmongkol, T.; Mulholland, A. J.; Hannongbua, S., High-Level QM/MM Calculations Support the Concerted Mechanism for Michael Addition and Covalent Complex Formation in Thymidylate Synthase. *Journal of Chemical Theory and Computation* **2015,** *11* (2), 713-722.

31. Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, T. F., Accurate and systematically improvable density functional theory embedding for correlated wavefunctions. *Journal of Chemical Physics* **2014,** *140* (18).

32. Bennie, S. J.; van der Kamp, M. W.; Pennifold, R. C. R.; Stella, M.; Manby, F. R.; Mulholland, A. J., A Projector-Embedding Approach for Multiscale Coupled-Cluster Calculations Applied to Citrate Synthase. *Journal of Chemical Theory and Computation* **2016,** *12* (6), 2689-2697.

33. Zhang, X. L.; Bennie, S. J.; van der Kamp, M. W.; Glowacki, D. R.; Manby, F. R.; Mulholland, A. J., Multiscale analysis of enantioselectivity in enzyme-catalysed 'lethal synthesis' using projector-based embedding. *Royal Society Open Science* **2018,** *5* (2), 171390.

34. Bennie, S. J.; Stella, M.; Miller, T. F.; Manby, F. R., Accelerating wavefunction in density-functional-theory embedding by truncating the active basis set. *Journal of Chemical Physics* **2015,** *143* (2), 024105-1 - 024105-7

35. Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, T. F., A Simple, Exact Density-Functional-Theory Embedding Scheme. *Journal of Chemical Theory and Computation* **2012,** *8* (8), 2564-2568.

36. Bennie, S. J.; Curchod, B. F. E.; Manby, F. R.; Glowacki, D. R., Pushing the Limits of EOM-CCSD with Projector-Based Embedding for Excitation Energies. *Journal of Physical Chemistry Letters* **2017,** *8* (22), 5559-5565.

37. Stella, M.; Bennie, S. J.; Manby, F. R., Computational study of adsorption of cobalt on benzene and coronene. *Molecular Physics* **2015,** *113* (13-14), 1858-1864.

38. Lyne, P. D.; Mulholland, A. J.; Richards, W. G., Insights into chorismate mutase catalysis from a combined QM/MM simulation of the enzyme reaction. *Journal of the American Chemical Society* **1995,** *117* (45), 11345-11350.

39. Ranaghan, K. E.; Mulholland, A. J., Investigations of enzyme-catalysed reactions with combined quantum mechanics/molecular mechanics (QM/MM) methods. *International Reviews in Physical Chemistry* **2010,** *29* (1), 65-133.

40. Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J., Transition structure selectivity in enzyme catalysis: a QM/MM study of chorismate mutase. *Theo. Chem. Acc.* **2001,** *105*, 207-212.

41. Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J., Conformational equilibrium of chorismate. A QM/MM theoretical study combining statistical simulations and geometry optimisations in gas phase and in aqueous solution. *J. Mol. Struct. (THEOCHEM)* **2003,** *632*, 197-206.

42. Marti, S.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J., Preorganization and reorganization as related factors in enzyme catalysis: The chorismate mutase case. *Chemistry-a European Journal* **2003,** *9* (4), 984-991.

43. Štrajbl, M.; Shurki, A.; Kato, M.; Warshel, A., Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilisation. *J. Am. Chem. Soc.* **2003,** *125*, 10228-10237.

44. Ghysels, A.; Woodcock, H. L.; Larkin, J. D.; Miller, B. T.; Shao, Y.; Kong, J.; Neck, D. V.; Speybroeck, V. V.; Waroquier, M.; Brooks, B. R., Efficient Calculation of QM/MM Frequencies with the Mobile Block Hessian. *Journal of Chemical Theory and Computation* **2011,** *7* (2), 496-514.

45. Woodcock, H. L.; Hodoscek, M.; Sherwood, P.; Lee, Y. S.; Schaefer, H. F.; Brooks, B. R., Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase. *Theoretical Chemistry Accounts* **2003,** *109* (3), 140-148.

46. Brickel, S.; Meuwly, M., Molecular Determinants for Rate Acceleration in the Claisen Rearrangement Reaction. *Journal of Physical Chemistry B* **2019,** *123* (2), 448-456.

47. Kast, P.; Asif-Ullah, M.; Hilvert, D., Is Chorismate Mutase a Prototypic Entropy Trap? *Tetrahedron Lett.* **1996,** *37* (16), 2691-2694.

48. Lever, G.; Cole, D. J.; Lonsdale, R.; Ranaghan, K. E.; Wales, D. J.; Mulholland, A. J.; Skylaris, C.-K.; Payne, M. C., Large-Scale Density Functional Theory Transition State Searching in Enzymes. *Journal of Physical Chemistry Letters* **2014,** *5* (21), 3614-3619.

49. Ranaghan, K.; Ridder, L.; Szefczyk, B.; Sokalski, W.; Hermann, J.; Mulholland, A., Insights into enzyme catalysis from QM/MM modelling: transition state stabilization in chorismate mutase. *Molecular Physics* **2003,** *101* (17), 2695-2714.

50. Guimaraes, C. R. W.; Repasky, M. P.; Chandrasekhar, J.; Tirado-Rives, J.; Jorgensen, W. L., Contributions of conformational compression and preferential transition state stabilization to the rate enhancement by chorismate mutase. *J. Am. Chem. Soc.* **2003,** *125* (23), 6892-6899.

51. Ranaghan, K. E.; Ridder, L.; Szefczyk, B.; Sokalski, W. A.; Hermann, J. C.; Mulholland, A. J., Transition state stabilization and substrate strain in enzyme catalysis: ab initio QM/MM modelling of the chorismate mutase reaction. *Org. Biomol. Chem.* **2004,** *2*, 968-980.

52. Pruitt, S. R.; Steinmann, C., Mapping Interaction Energies in Chorismate Mutase with the Fragment Molecular Orbital Method. *Journal of Physical Chemistry A* **2017,** *121* (8), 1798-1808.

53. Claeyssens, F.; Ranaghan, K.; Manby, F.; Harvey, J.; Mulholland, A., Multiple high-level QM/MM reaction paths demonstrate transition-state stabilization in chorismate mutase: correlation of barrier height with transition-state stabilization. *Chemical Communications* **2005,** (40), 5068-5070.

54. Claeyssens, F.; Ranaghan, K. E.; Lawan, N.; Macrae, S. J.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J., Analysis of chorismate mutase catalysis by QM/MM modelling of enzyme-catalysed and uncatalysed reactions. *Organic & Biomolecular Chemistry* **2011,** *9* (5), 1578-1590.

55. Kienhofer, A.; Kast, P.; Hilvert, D., Selective stabilization of the chorismate mutase transition state by a positively charged hydrogen bond donor. *Journal of the American Chemical Society* **2003,** *125* (11), 3206-3207.

56. Burschowsky, D.; van Eerde, A.; Okvist, M.; Kienhofer, A.; Kast, P.; Hilvert, D.; Krengel, U., Electrostatic transition state stabilization rather than reactant destabilization provides the chemical basis for efficient chorismate mutase catalysis. *Proceedings of the National Academy of Sciences of the United States of America* **2014,** *111* (49), 17516-17521.

57. Guimaraes, C. R. W.; Udier-Blagovic, M.; Tubert-Brohman, I.; Jorgensen, W. L., Effects of Arg90 neutralization on the enzyme-catalyzed rearrangement of chorismate to prephenate. *J. Chem. Theory and Comput.* **2005,** *1* (4), 617-625.

58. Burschowsky, D.; Krengel, U.; Uggerud, E.; Balcells, D., Quantum chemical modeling of the reaction path of chorismate mutase based on the experimental substrate/product complex. *Febs Open Bio* **2017,** *7* (6), 789-797.

59. Chook, Y. M.; Ke, H. M.; Lipscomb, W. N., Crystal-Structures of the Monofunctional Chorismate Mutase from Bacillus-Subtilis and its Complex with a Transition-State Analog. *Proceedings of the National Academy of Sciences of the United States of America* **1993,** *90* (18), 8600-8603.

60. Bartlett, P. A.; Johnson, C. R., An Inhibitor of Chorismate Mutase Resembling the Transition-State Conformation. *Journal of the American Chemical Society* **1985,** *107* (25), 7792-7793.

61. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., CHARMM - A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* **1983,** *4* (2), 187-217.

62. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.

63. Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S., A self-consistent charge density-functional based tight-binding scheme for large biomolecules. *Physica Status Solidi B-Basic Research* **2000,** *217* (1), 357-376.

64. Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M., A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *Journal of Physical Chemistry B* **2001,** *105* (2), 569-585.

65. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.;

Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *Journal of Physical Chemistry B* **1998,** *102* (18), 3586-3616.

66. Henkelman, G.; Uberuaga, B. P.; Jonsson, H., A Climbing Image Nudged Elastic Band Method For Finding Saddle Points and Minimum Energy Paths. *J. Chem. Phys.* **2000,** *113* (22), 9901-9904.

67. Sheppard, D.; Terrell, R.; Henkelman, G., Optimization Methods for Finding Minimum Energy Paths. *J. Chem. Phys.* **2008,** *128* (13).

68. Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A., Jaguar: A High-Performance Quantum Chemistry Software Program with Strengths in Life and Materials Sciences. *International Journal of Quantum Chemistry* **2013,** *113* (18), 2110-2142.

69. Ren, P.; Ponder, J. W., Tinker Polarizable Atomic Multipole Force Field for Proteins. *Abstracts of Papers of the American Chemical Society* **2002,** *224*, U473-U473.

70. Harvey, J. N., Spin-Forbidden CO Ligand Recombination in Myoglobin. *Faraday Discussions* **2004,** *127*, 165-177.

71. Foloppe, N.; MacKerell, A. D., All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal of Computational Chemistry* **2000,** *21* (2), 86-104.

72. Mulliken, R. S., Electronic Population Analysis on LCAO-MO Molecular Wave Functions .1. *Journal of Chemical Physics* **1955,** *23* (10), 1833-1840.

73. Zhao, Y.; Truhlar, D. G., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts* **2008,** *120* (1-3), 215-241.

74. Becke, A. D., A New Mixing of Hartree-Fock and Local Density-Functional Theories. *Journal of Chemical Physics* **1993,** *98* (2), 1372-1377.

75. Becke, A. D., Density-Functional Thermochemistry .3. The Role of Exact Exchange. *Journal of Chemical Physics* **1993,** *98* (7), 5648-5652.

76. Zhao, Y.; Truhlar, D. G., A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *Journal of Chemical Physics* **2006,** *125* (19), 194101-1 - 194101-18.

77. Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized gradient approximation made simple. *Physical Review Letters* **1996,** *77* (18), 3865-3868.

78. Altun, A.; Shaik, S.; Thiel, W., Systematic QM/MM investigation of factors that affect the cytochrome P450-catalyzed hydrogen abstraction of camphor. *Journal of Computational Chemistry* **2006,** *27* (12), 1324-1337.

79. Vasilevskaya, T.; Khrenova, M. G.; Nemukhin, A. V.; Thiel, W., Methodological aspects of QM/MM calculations: A case study on matrix metalloproteinase-2. *Journal of Computational Chemistry* **2016,** *37* (19), 1801-1809.

80. Dixit, M.; Weitman, M.; Gao, J. L.; Major, D. T., Chemical Control in the Battle against Fidelity in Promiscuous Natural Product Biosynthesis: The Case of Trichodiene Synthase. *Acs Catalysis* **2017,** *7* (1), 812-818.

81. O'Brien, T. E.; Bertolani, S. J.; Tantillo, D. J.; Siegel, J. B., Mechanistically informed predictions of binding modes for carbocation intermediates of a sesquiterpene synthase reaction. *Chemical Science* **2016,** *7* (7), 4009-4015.

82. Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklaß, A.; O'Neill, D. P.; Palmieri, P.; Peng, D.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M. *Molpro: a general purpose quantum chemistry program package*, Cardiff, UK, 2014.

83. Dunning, T. H., Gaussian-Basis Sets for Use in Correlated Molecular Calculations .1. The Atoms Boron through Neon and Hydrogen. *Journal of Chemical Physics* **1989,** *90* (2), 1007-1023.

84. Kendall, R. A.; Dunning, T. H.; Harrison, R. J., Electron-Affinities of the 1st-Row Atoms Revisited - Systematic Basis-Sets and Wavefunctions. *Journal of Chemical Physics* **1992,** *96* (9), 6796-6806.

85. Woon, D. E.; Dunning, T. H., Gaussian-Basis Sets for Use in Correlated Molecular Calculations .3. The Atoms Aluminium through Argon. *Journal of Chemical Physics* **1993,** *98* (2), 1358-1371.

86. Moller, C.; Plesset, M. S., Note on an approximation treatment for many-electron systems. *Physical Review* **1934,** *46* (7), 0618-0622.

87. Grimme, S., Improved second-order Moller-Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies. *Journal of Chemical Physics* **2003,** *118* (20), 9095-9102.

88. Bartlett, R. J.; Musial, M., Coupled-cluster theory in quantum chemistry. *Reviews of Modern Physics* **2007,** *79* (1), 291-352.

89. Bartlett, R. J., Coupled-cluster approach to molecular structure and spectra: A step toward predictive quantum chemistry. *Journal of Physical Chemistry* **1989,** *93* (5), 1697-1708.

90. Bartlett, R. J.; Watts, J. D.; Kucharski, S. A.; Noga, J., NONITERATIVE 5TH-ORDER TRIPLE AND QUADRUPLE EXCITATION-ENERGY CORRECTIONS IN CORRELATED METHODS. *Chemical Physics Letters* **1990,** *165* (6), 513-522.

91. Mata, R. A.; Werner, H. J.; Schutz, M., Correlation regions within a localized molecular orbital approach. *Journal of Chemical Physics* **2008,** *128* (14), 144106-1 - 144106-8.

92. Neese, F.; Hansen, A.; Liakos, D. G., Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis. *Journal of Chemical Physics* **2009,** *131* (6), 064103-1 - 064103-14.

93. Yang, J.; Chan, G. K. L.; Manby, F. R.; Schutz, M.; Werner, H. J., The orbital-specific-virtual local coupled cluster singles and doubles method. *Journal of Chemical Physics* **2012,** *136* (14), 144105-1 - 144105-14.

94. Hampel, C.; Werner, H. J., Local treatment of electron correlation in coupled cluster theory. *Journal of Chemical Physics* **1996,** *104* (16), 6286-6297.

95. Kaliman, I. A.; Krylov, A. I., New Algorithm for Tensor Contractions on Multi-Core CPUs, GPUs, and Accelerators Enables CCSD and EOM-CCSD Calculations with over 1000 Basis Functions on a Single Compute Node. *Journal of Computational Chemistry* **2017,** *38* (11), 842-853.

96. Hohenstein, E. G.; Parrish, R. M.; Martinez, T. J., Tensor hypercontraction density fitting. I. Quartic scaling second- and third-order Moller-Plesset perturbation theory. *Journal of Chemical Physics* **2012,** *137* (4), 044103-1 - 044103-10.

97. Masgrau, L.; Roujeinikova, A.; Johannissen, L.; Hothi, P.; Basran, J.; Ranaghan, K.; Mulholland, A.; Sutcliffe, M.; Scrutton, N.; Leys, D., Atomic description of an enzyme reaction dominated by proton tunneling. *Science* **2006,** *312* (5771), 237-241.

98. Ranaghan, K. E.; Mulholland, A. J., Computer Simulations of Quantum Tunnelling in Enzyme-Catalysed Hydrogen Transfer Reactions. *Interdisciplinary Sciences-Computational Life Sciences* **2010,** *2* (1), 78-97.

99. Ranaghan, K. E.; Ridder, L.; Szefczyk, B.; Sokalski, W. A.; Hermann, J. C.; Mulholland, A. J., Insights into enzyme catalysis for QM/MM modelling:

transition state stabilisation in chorismate mutase. *Mol. Phys.* **2003,** *101*, 2695-2714.

100. Lodola, A.; Sirirak, J.; Fey, N.; Rivara, S.; Mor, M.; Mulholland, A. J., Structural Fluctuations in Enzyme-Catalyzed Reactions: Determinants of Reactivity in Fatty Acid Amide Hydrolase from Multivariate Statistical Analysis of Quantum Mechanics/Molecular Mechanics Paths. *Journal of Chemical Theory and Computation* **2010,** *6* (9), 2948-2960.

101. Lonsdale, R.; Houghton, K. T.; Zurek, J.; Bathelt, C. M.; Foloppe, N.; de Groot, M. J.; Harvey, J. N.; Mulholland, A. J., Quantum Mechanics/Molecular Mechanics Modeling of Regioselectivity of Drug Metabolism in Cytochrome P450 2C9. *Journal of the American Chemical Society* **2013,** *135* (21), 8001-8015.

102. Liu, Z.; Patel, C.; Harvey, J. N.; Sunoj, R. B., Mechanism and reactivity in the Morita-Baylis-Hillman reaction: the challenge of accurate computations. *Physical Chemistry Chemical Physics* **2017,** *19* (45), 30647-30657.

103. van der Kamp, M. W.; McGeagh, J. D.; Mulholland, A. J., "Lethal Synthesis" of Fluorocitrate by Citrate Synthase Explained through QM/MM Modeling. *Angewandte Chemie-International Edition* **2011,** *50* (44), 10349-10351.

104. Lonsdale, R.; Harvey, J. N.; Mulholland, A. J., Compound I Reactivity Defines Alkene Oxidation Selectivity in Cytochrome P450cam. *Journal of Physical Chemistry B* **2010,** *114* (2), 1156-1162.

105. Ryde, U., How Many Conformations Need To Be Sampled To Obtain Converged QM/MM Energies? The Curse of Exponential Averaging. *Journal of Chemical Theory and Computation* **2017,** *13* (11), 5745-5752.

106. Bowman, A. L.; Ridder, L.; Rietjens, I.; Vervoort, J.; Mulholland, A. J., Molecular determinants of xenobiotic metabolism: QM/MM simulation of the conversion of 1-chloro-2,4-dinitrobenzene catalyzed by M1-1 glutathione S-transferase. *Biochemistry* **2007,** *46* (21), 6353-6363.

107. Ridder, L.; Rietjens, I.; Vervoort, J.; Mulholland, A. J., Quantum mechanical/molecular mechanical free energy Simulations of the glutathione S-transferase (M1-1) reaction with phenanthrene 9,10-oxide. *Journal of the American Chemical Society* **2002,** *124* (33), 9926-9936.

108. Kamerlin, S. C. L.; Warshel, A., The EVB as a quantitative tool for formulating simulations and analyzing biological and chemical reactions. *Faraday Discussions* **2010,** *145*, 71-106.

109. Szefczyk, B.; Sokalski, W. A.; Ranaghan, K. E.; Mulholland, A. J., Differential transition-state stabilization in enzyme catalysis :Quantum chemical analysis of interactions in the chorismate mutase reaction and prediction of the optimal catalytic field. *J. Am. Chem. Soc.* **2004,** *126* (49), 16148-16159.

110. Rinaldi, S.; Van der Kamp, M. W.; Ranaghan, K. E.; Mulholland, A. J.; Colombo, G., Understanding Complex Mechanisms of Enzyme Reactivity: The Case of Limonene-1,2-Epoxide Hydrolases. *Acs Catalysis* **2018,** *8* (7), 5698-5707.

111. Harvey, J. N.; Aggarwal, V. K.; Bathelt, C. M.; Carreon-Macedo, J. L.; Gallagher, T.; Holzmann, N.; Mulholland, A. J.; Robiette, R., QM and QM/MM studies of selectivity in organic and bioorganic chemistry. *Journal of Physical Organic Chemistry* **2006,** *19* (8-9), 608-615.

112. Lonsdale, R.; Harvey, J. N.; Mulholland, A. J., A practical guide to modelling enzyme-catalysed reactions. *Chemical Society Reviews* **2012,** *41* (8), 3025-3038.

113. Callegari, D.; Ranaghan, K. E.; Woods, C. J.; Minari, R.; Tiseo, M.; Mor, M.; Mulholland, A. J.; Lodola, A., L718Q mutant EGFR escapes covalent inhibition by stabilizing a non-reactive conformation of the lung cancer drug osimertinib. *Chemical Science* **2018,** *9* (10), 2740-2749.

114. Lodola, A.; Mor, M.; Rivara, S.; Christov, C.; Tarzia, G.; Piomelli, D.; Mulholland, A. J., Identification of productive inhibitor binding orientation in fatty acid amide hydrolase (FAAH) by QM/MM mechanistic modelling. *Chemical Communications* **2008,** (2), 214-216.

115. Senn, H. M.; Thiel, W., QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Edit.* **2009,** *48* (7), 1198-1229.

116. Lonsdale, R.; Ranaghan, K. E.; Mulholland, A. J., Computational enzymology. *Chemical Communications* **2010,** *46* (14), 2354-2372.

117. Huggins, D. J.; Biggin, P. C.; Dämgen, M. A.; Essex, J. W.; Harris, S. A.; Henchman, R. H.; Khalid, S.; Kuzmanic, A.; Laughton, C. A.; Michel, J.; Mulholland, A. J.; Rosta, E.; Sansom, M. S. P.; van der Kamp, M. W., Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity. *WIREs Comput Mol Sci* **2018**.

118. Sousa, S. F.; Ribeiro, A. J. M.; Neves, R. P. P.; Bras, N. F.; Cerqueira, N.; Fernandes, P. A.; Ramos, M. J., Application of quantum mechanics/molecular mechanics methods in the study of enzymatic reaction mechanisms. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2017,** *7* (2).

119. Ranaghan, K. E.; Hung, J. E.; Bartlett, G. J.; Mooibroek, T. J.; Harvey, J. N.; Woolfson, D. N.; van der Donk, W. A.; Mulholland, A. J., A catalytic role for methionine revealed by a combination of computation and experiments on phosphite dehydrogenase. *Chemical Science* **2014,** *5* (6), 2191-2199.

120. Ridder, L.; Mulholland, A. J.; Rietjens, I.; Vervoort, J., A quantum mechanical/molecular mechanical study of the hydroxylation of phenol and halogenated derivatives by phenol hydroxylase. *Journal of the American Chemical Society* **2000,** *122* (36), 8728-8738.

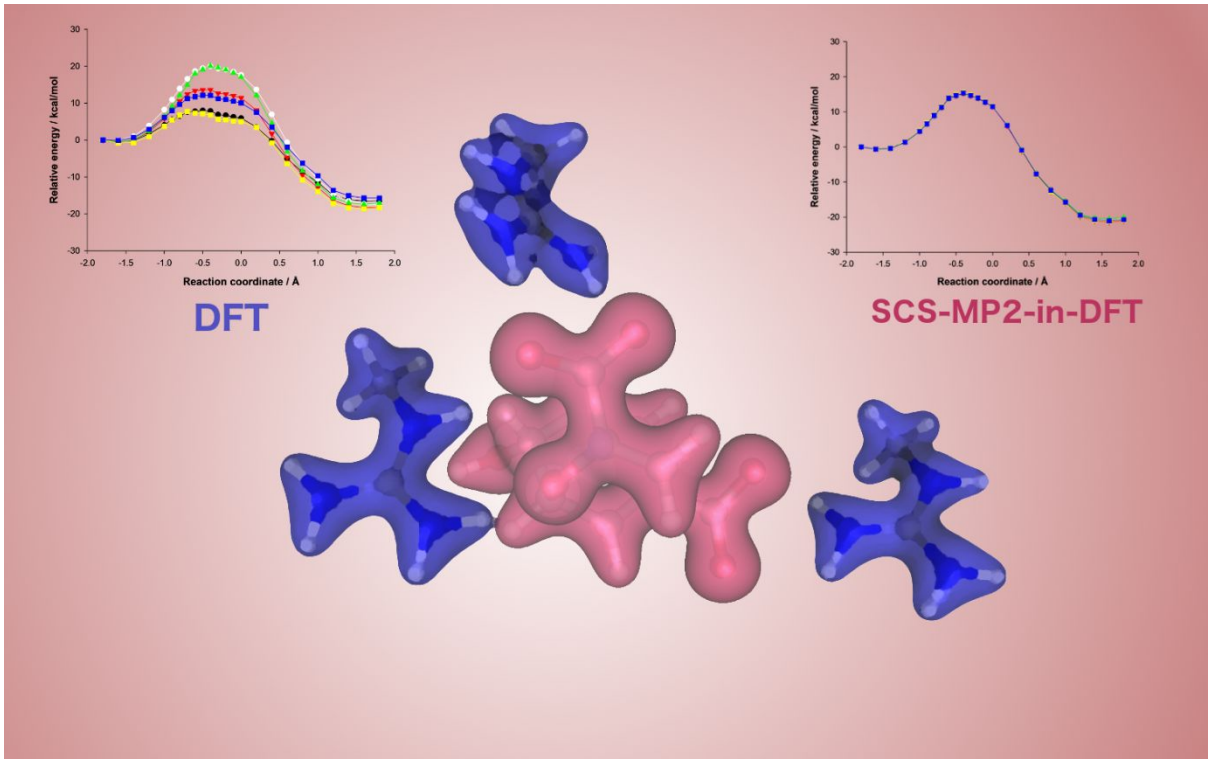
121. Lonsdale, R.; Harvey, J. N.; Mulholland, A. J., Inclusion of Dispersion Effects Significantly Improves Accuracy of Calculated Reaction Barriers for Cytochrome P450 Catalyzed Reactions. *Journal of Physical Chemistry Letters* **2010,** *1* (21), 3232-3237.

122. Grimme, S., Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry* **2004,** *25* (12), 1463-1473.

123. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *Journal of Chemical Physics* **2010,** *132* (15).

124. Mulholland, A. J., Chemical accuracy in QM/MM calculations on enzyme-catalysed reactions. *Chemistry Central Journal* **2007,** *1*.

TOC Graphic


Relative energy / kcal/mol
Reaction coordinate / Å
DFT
Relative energy / kcal/mol
Reaction coordinate / Å
SCS-MP2-in-DFT